



EMR Amazon-Veröffentlichungsleitfaden

Amazon EMR



Amazon EMR: EMRAmazon-Veröffentlichungsleitfaden

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

Die Handelsmarken und Handelsaufmachung von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, durch die Kunden irregeführt werden könnten oder Amazon in schlechtem Licht dargestellt oder diskreditiert werden könnte. Alle anderen Marken, die nicht im Besitz von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise mit Amazon verbunden sind oder von Amazon gesponsert werden.

Table of Contents

Über Amazon EMR Releases	1
Standardunterstützung	2
EMRAmazon-Releases verstehen	2
Versionierung der Versionen	3
Support-Richtlinie	4
Überlegungen	7
Release-Versionen von Amazon EMR 7.x	9
Anwendungsversionen in Amazon EMR 7.x-Versionen	9
emr-7.2.0	10
emr-7.1.0	66
emr-7.0.0	94
Release-Versionen von Amazon EMR 6.x	132
Anwendungsversionen in Amazon-EMR-6.x-Versionen	133
emr-6.15.0	134
emr-6.14.0	182
emr-6.13.0	230
emr-6.12.0	286
emr-6.11.1	346
emr-6.11.0	403
emr-6.10.1	441
emr-6.10.0	498
emr-6.9.1	539
emr-6.9.0	597
emr-6.8.1	645
emr-6.8.0	702
emr-6.7.0	750
emr-6.6.0	820
emr-6.5.0	893
emr-6.4.0	921
emr-6.3.1	955
emr-6.3.0	982
emr-6.2.1	1016
emr-6.2.0	1043
emr-6.1.1	1077

emr-6.1.0	1098
emr-6.0.1	1124
emr-6.0.0	1142
Amazon EMR 5.x-Release-Versionen	1164
Anwendungsversionen in Amazon-EMR-5.x-Versionen (PNG)	1168
emr-5.36.2	1168
emr-5.36.1	1196
emr-5.36.0	1252
emr-5.35.0	1290
emr-5.34.0	1318
emr-5.33.1	1343
emr-5.33.0	1373
emr-5.32.1	1395
emr-5.32.0	1420
emr-5.31.1	1449
emr-5.31.0	1469
emr-5.30.2	1492
emr-5.30.1	1512
emr-5.30.0	1534
emr-5.29.0	1558
emr-5.28.1	1577
emr-5.28.0	1596
emr-5.27.1	1616
emr-5.27.0	1633
emr-5.26.0	1653
emr-5.25.0	1673
emr-5.24.1	1692
emr-5.24.0	1710
emr-5.23.1	1729
emr-5.23.0	1745
emr-5.22.0	1764
emr-5.21.2	1783
emr-5.21.1	1800
emr-5.21.0	1817
emr-5.20.1	1836
emr-5.20.0	1853

emr-5.19.1	1873
emr-5.19.0	1890
emr-5.18.1	1908
emr-5.18.0	1924
emr-5.17.2	1941
emr-5.17.1	1958
emr-5.17.0	1974
emr-5.16.1	1992
emr-5.16.0	2008
emr-5.15.1	2025
emr-5.15.0	2041
emr-5.14.2	2057
emr-5.14.1	2074
emr-5.14.0	2090
emr-5.13.1	2108
emr-5.13.0	2123
emr-5.12.3	2139
emr-5.12.2	2155
emr-5.12.1	2170
emr-5.12.0	2186
emr-5.11.4	2203
emr-5.11.3	2218
emr-5.11.2	2234
emr-5.11.1	2249
emr-5.11.0	2265
emr-5.10.1	2281
emr-5.10.0	2296
emr-5.9.1	2313
emr-5.9.0	2328
emr-5.8.3	2345
emr-5.8.2	2359
emr-5.8.1	2374
emr-5.8.0	2389
emr-5.7.1	2406
emr-5.7.0	2421
emr-5.6.1	2436

emr-5.6.0	2451
emr-5.5.4	2466
emr-5.5.3	2481
emr-5.5.2	2496
emr-5.5.1	2510
emr-5.5.0	2525
emr-5.4.1	2541
emr-5.4.0	2556
emr-5.3.2	2571
emr-5.3.1	2586
emr-5.3.0	2601
emr-5.2.3	2616
emr-5.2.2	2630
emr-5.2.1	2645
emr-5.2.0	2660
emr-5.1.1	2675
emr-5.1.0	2690
emr-5.0.3	2705
emr-5.0.2	2719
emr-5.0.1	2733
emr-5.0.0	2747
Amazon EMR 4.x-Release-Versionen	2762
Anwendungsversionen in Amazon-EMR-4.x-Versionen	2764
Versionsunterschiede	2764
emr-4.9.6	2812
emr-4.9.5	2827
emr-4.9.4	2841
emr-4.9.3	2855
emr-4.9.2	2869
emr-4.9.1	2883
emr-4.8.5	2897
emr-4.8.4	2912
emr-4.8.3	2926
emr-4.8.2	2940
emr-4.8.1	2955
emr-4.8.0	2969

emr-4.7.4	2983
emr-4.7.3	2997
emr-4.7.2	3010
emr-4.7.1	3024
emr-4.7.0	3037
emr-4.6.1	3051
emr-4.6.0	3064
emr-4.5.0	3077
emr-4.4.0	3088
emr-4.3.0	3100
emr-4.2.0	3111
emr-4.1.0	3122
emr-4.0.0	3130
Versionen 2.x und 3.x AMI	3138
Erstellen eines Clusters	3139
Installieren von Anwendungen	3141
Anpassen von Konfigurationen	3141
Hive	3149
HBase	3159
Pig	3173
Spark	3179
S3 DistCp	3183
Neuerungen?	3186
Amazon EMR 7.2.0	3186
Amazon EMR 6,15.0	3189
Amazon EMR 5,36,2	3208
SigV4-Kompatibilität	3212
Ansatz zur Minderung CVE -2021-44228	3213
Amazon EMR Bootstrap-Aktionslösung für Log4j CVE -2021-44228 & -2021-45046 CVE ..	3214
Häufig gestellte Fragen	3221
Archiv	3224
Version 6.14.0	3224
Version 6.13.0	3244
Version 6.12.0	3270
Version 6.11.1	3301
Version 6.11.0	3328

Version 6.10.0	3340
Version 6.9.0	3354
Version 6.8.0	3374
Version 6.7.0	3395
Version 6.6.0	3439
Version 5.35.0	3485
Version 5.34.0	3489
Version 6.5.0	3492
Version 6.4.0	3494
Version 5.32.0	3502
Version 6.2.0	3508
Version 5.31.0	3516
Version 6.1.0	3522
Version 6.0.0	3528
Version 5.30.1	3534
Version 5.30.0	3539
Version 5.29.0	3544
Version 5.28.1	3546
Version 5.28.0	3547
Veröffentlichung 5.27.0	3549
Version 5.26.0	3552
Version 5.25.0	3555
Version 5.24.1	3558
Version 5.24.0	3559
Version 5.23.0	3561
Version 5.22.0	3564
Version 5.21.1	3567
Version 5.21.0	3568
Version 5.20.0	3571
Version 5.19.0	3574
Version 5.18.0	3576
Version 5.17.1	3577
Version 5.17.0	3577
Version 5.16.0	3579
Version 5.15.0	3580
Version 5.14.1	3581

Version 5.14.0	3581
Version 5.13.0	3584
Version 5.12.2	3584
Version 5.12.1	3585
Version 5.12.0	3585
Version 5.11.3	3587
Version 5.11.2	3587
Version 5.11.1	3587
Version 5.11.0	3588
Version 5.10.0	3589
Version 5.9.0	3591
Version 5.8.2	3593
Version 5.8.1	3594
Version 5.8.0	3594
Version 5.7.0	3596
Version 5.6.0	3597
Version 5.5.3	3598
Version 5.5.2	3598
Version 5.5.1	3598
Version 5.5.0	3599
Version 5.4.0	3600
Version 5.3.1	3601
Version 5.3.0	3601
Version 5.2.2	3602
Version 5.2.1	3602
Version 5.2.0	3603
Version 5.1.0	3604
Version 5.0.3	3605
Version 5.0.0	3605
Version 4.9.5	3607
Version 4.9.4	3608
Version 4.9.3	3608
Version 4.9.2	3608
Version 4.9.1	3608
Version 4.8.4	3609
Version 4.8.3	3609

Version 4.8.2	3610
Version 4.8.0	3611
Version 4.7.2	3612
Version 4.7.1	3612
Version 4.7.0	3613
Version 4.6.0	3615
Version 4.5.0	3616
Version 4.4.0	3617
Version 4.3.0	3619
Version 4.2.0	3621
Anwendungen konfigurieren	3623
Anwendungen bei der Clustererstellung konfigurieren	3626
Geben Sie beim Erstellen eines Clusters eine Konfiguration in der Konsole an	3626
Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie	3626
Geben Sie eine Konfiguration mit Java anSDK, wenn Sie einen Cluster erstellen	3627
Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren	3628
Überlegungen bei der Neukonfiguration einer Instance-Gruppe	3629
Eine Instance-Gruppe in der Konsole neu konfigurieren	3632
Rekonfigurieren Sie eine Instanzgruppe mit dem CLI	3633
Rekonfigurieren Sie eine Instanzgruppe mithilfe von Java SDK	3638
Fehlerbehebung	3639
Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern	3642
Ein Secret erstellen	3642
Gewähren EMR Sie Amazon Zugriff, um das Geheimnis abzurufen	3643
Das Secret in einer Konfigurationsklassifizierung verwenden	3643
Den Secret-Wert aktualisieren	3645
Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren	3645
Überlegungen	3645
Überschreiben Sie die JVM	3648
Service-Ports	3650
Anwendungsbenutzer	3652
Überprüfen von Abhängigkeiten mit dem Artefakt-Repository	3654
EMRDateisystem (EMRFS)	3657
Consistent view	3659
Die konsistente Ansicht aktivieren	3663
Verstehen, wie EMRFS Consistent View Objekte in Amazon S3 verfolgt	3665

Logik für Wiederholversuche	3666
EMRFSkonsistente Metadaten anzeigen	3668
Konsistenzbenachrichtigungen für CloudWatch und Amazon konfigurieren SQS	3671
Konfigurieren der konsistenten Ansicht	3673
EMRFSCLIBefehlsreferenz	3679
Autorisieren des Zugriffs auf EMRFS Daten in Amazon S3	3690
Einen Anbieter für benutzerdefinierte Anmeldeinformationen für EMRFS Daten in Amazon S3 erstellen	3691
Den AWS Security Token Service Standardendpunkt verwalten	3692
Amazon S3 S3-Verschlüsselung mithilfe von EMRFS Eigenschaften angeben	3694
Wird für die Verschlüsselung verwendet AWS KMS keys EMRFS	3695
Serverseitige Verschlüsselung im Amazon S3	3696
Clientseitige Verschlüsselung für Amazon S3	3699
CloudWatch Amazon-Vertreter	3708
Erstellen eines -Clusters	3710
Erforderliche Berechtigungen	3710
Erforderlicher Endpunkt	3711
Den Cluster erstellen	3711
Standardmetriken	3712
Konfiguration	3715
Amazon EMR 7.1.0	3715
Amazon EMR 7.0.0	3750
Überlegungen	3756
Verlauf	3756
Delta Lake	3758
Einführung	3759
Delta-Lake-Cluster verwenden	3759
Delta Lake mit Flink	3759
Delta Lake mit Trino	3764
Delta Lake mit Spark	3766
Delta Lake mit Spark und Glue	3772
Überlegungen	3773
Verlauf	3773
Flink	3775
Erstellen eines Clusters mit Flink	3778
Flink konfigurieren	3779

Hive und Glue	3779
Konfigurationsdatei	3781
Mehrere Primärknoten	3783
Speicherprozessgröße	3784
Protokollausgabedateigröße	3784
Java 11	3786
Flink-Aufträge	3790
Eine Flink-YARN-Anwendung als Schritt auf einem Cluster mit langer Laufzeit starten	3791
Senden Sie Ihre Arbeit an eine bestehende Flink-Anwendung auf einem Cluster mit langer Laufzeit	3792
Senden eines kurzlebigen Flink-Auftrags	3794
Flink-Scala-Shell	3796
Flink-Benutzeroberfläche	3797
Flink Autoscaler	3798
Übersicht	3798
Überlegungen	3798
Autoscaler aktivieren	3799
Konfigurationen	3800
Neustartzeiten optimieren	3806
Aufgabenlokale Wiederherstellung	3806
Inkrementelle Prüfpunkte	3807
Differenzierte Wiederherstellung	3808
Kombinierter Neustart	3809
Flink mit Zeppelin	3810
Einführung	3810
Voraussetzungen	3811
Zeppelin-Flink auf einem EMR-Cluster konfigurieren	3811
Führen Sie Flink-Aufträge mit Zeppelin-Flink auf einem EMR-Cluster aus	3813
Versionsverlauf von Flink	3818
Versionshinweise zu Flink nach Version	3856
Ganglia	3857
Erstellen eines Clusters mit Ganglia	3859
Anzeigen von Ganglia-Metriken	3860
Hadoop- und Spark-Metriken in Ganglia	3861
Versionsverlauf von Ganglia	3862
Hadoop	3908

Konfigurieren von Hadoop	3910
Aufgabenkonfiguration	3911
Hadoop-Daemon-Konfigurationseinstellungen	4271
HDFSKonfiguration	4541
Transparente Verschlüsselung HDFS bei Amazon EMR	4543
Konfiguration HDFS transparenter Verschlüsselung	4544
Überlegungen zur HDFS transparenten Verschlüsselung	4547
Hadoop-Schlüsselverwaltungsserver (KMS)	4547
HDFStransparente Verschlüsselung auf EMR Clustern mit mehreren Master-Knoten	4551
Erstellen oder Ausführen einer Hadoop-Anwendung	4553
Erstellen von Binärdateien mit Amazon EMR	4553
Daten per Streaming verarbeiten	4556
Verarbeiten Sie Daten mit einem benutzerdefinierten JAR	4561
Lesen Sie wiederhergestellte Objekte	4565
Beispiele	4565
Überlegungen	4566
Aktivieren Sie die Erkennung ungleichmäßiger Speicherzugriffe für Container YARN	4567
Hadoop-Versionsverlauf	4569
Hadoop-Versionshinweise nach Version	4631
HBase	4637
Einen Cluster erstellen mit HBase	4641
Einen Cluster HBase mithilfe der Konsole erstellen	4641
Erstellen eines Clusters HBase mit der AWS CLI	4641
HBaseauf Amazon S3 (Amazon S3 S3-Speichermodus)	4643
Aktivierung HBase auf Amazon S3	4644
Verwenden eines Lesereplikat-Clusters	4645
HFilePersistentes Tracking	4647
Betriebliche Überlegungen	4649
Write-Ahead-Logs (WAL) für Amazon EMR	4653
WALArbeitsbereiche	4655
Erforderliche Berechtigungen	4655
Aktivieren WAL	4656
Wiederherstellung von WAL	4659
Sicherheitskonfigurationen	4660
Verwenden AWS PrivateLink	4661
WALPreisgestaltung und Kennzahlen	4662

Arbeitsbereiche taggen WAL	4664
Überlegungen und Verfügbarkeit	4665
EMRWALCLIREferenz	4666
Verwenden der HBase Shell	4670
Erstellen einer Tabelle	4670
Eingeben eines Werts	4670
Abrufen eines Werts	4671
Löschen einer Tabelle	4671
Greifen Sie mit Hive auf HBase Tabellen zu	4671
Schnappschüsse verwenden HBase	4673
Erstellen eines Snapshots mit einer Tabelle	4673
Löschen eines Snapshots	4674
Anzeigen von Snapshot-Informationen	4674
Einen Snapshot nach Amazon S3 exportieren	4674
Importieren eines Snapshots von Amazon S3	4675
Eine Tabelle aus Snapshots in der HBase Shell wiederherstellen	4676
Konfigurieren HBase	4677
Änderungen an der Speicherzuweisung in YARN	4678
HBasePortnummern	4678
HBaseSeiteneinstellungen zur Optimierung	4679
Sehen Sie sich die HBase Benutzeroberfläche an	4681
Anzeige von HBase-Protokolldateien	4683
Monitor HBase mit Ganglia	4684
Migration von früheren Versionen HBase	4686
HBaseVersionsverlauf	4686
HCatalog	4759
Einen Cluster erstellen mit HCatalog	4761
Verwenden HCatalog	4762
Deaktivieren Sie direktes Schreiben bei der Verwendung von HCatalog HStorer	4762
Erstellen Sie eine Tabelle mit den HCat CLI und verwenden Sie diese Daten in Pig	4762
Zugreifen auf die Tabelle mit Spark SQL	4763
Beispiel: Erstellen Sie eine HCatalog Tabelle und schreiben Sie mit Pig in sie	4765
HCatalogVersionsverlauf	4766
Hive	4835
Unterschiede und Überlegungen zu Hive bei Amazon EMR	4838
Unterschiede zwischen Apache Hive auf Amazon EMR und Apache Hive	4838

Unterschiede in Hive zwischen EMR Amazon-Release-Version 4.x und 5.x	4839
Zusätzliche Funktionen von Hive bei Amazon EMR	4840
Einen externen Metastore für Hive konfigurieren	4846
Den AWS Glue-Datenkatalog als Metastore für Hive verwenden	4846
Verwenden einer externen My SQL Database oder Amazon Aurora	4854
Den Hive-JDBC-Treiber verwenden	4856
Verbessern der Hive-Leistung	4859
Aktivierung des optimierten Hive EMRFS S3-Committers	4859
Verwendung von S3 Select	4860
MSCKOptimierung	4863
Verwenden Sie Hive LLAP	4864
Aktivieren von LLAP	4864
Starten Sie LLAP auf Ihrem Cluster	4866
Überprüfen Sie den Status LLAP	4866
Startet oder stoppt LLAP	4867
Ändern Sie LLAP die Größe der Anzahl der Daemons	4867
Verschlüsselung im Hive	4867
Modulare Parquet-Verschlüsselung in Hive	4867
Verschlüsselung bei der Übertragung in HS2	4871
Versionsverlauf von Hive	4873
Hive-Versionshinweise nach Version	4943
Hudi	5027
Funktionsweise von Hudi	5029
Grundlegendes zu den Speichertypen von Datensätzen: „Beim Schreiben kopieren“ gegenüber „beim Lesen zusammenführen“	5030
Registrieren eines Hudi-Datensatzes bei Ihrem Metastore	5031
Überlegungen und Einschränkungen	5031
Erstellen eines Cluster mit installierter Hudi-Anwendung	5033
Arbeiten mit einem Hudi-Datensatz	5034
Initialisieren Sie eine Spark-Sitzung für Hudi	5038
Schreiben Sie in einen Hudi-Datensatz	5038
Upsert Daten	5044
Einen Datensatz löschen	5045
Aus einem Hudi-Datensatz lesen	5046
Benutze den Hudi CLI	5048
Hudi-Versionsverlauf	5049

Hue	5053
Hue-Versionen	5053
Unterstützte und nicht unterstützte Funktionen von Hue bei Amazon EMR	5056
Überlegungen	5057
Leistung bei großen Hue-Metadatatabellen	5057
Inkompatibilität zwischen Hue-Versionen	5058
Verbindung zur Hue-Webbenutzeroberfläche herstellen	5058
Hue mit einer Remote-Datenbank in Amazon verwenden RDS	5059
Fehlerbehebung	5061
Erweiterte Konfigurationen für Hue	5062
Konfigurieren Sie Hue für Benutzer LDAP	5062
Hue-Versionsverlauf	5066
Iceberg	5128
Funktionsweise von Iceberg	5129
Einen Cluster mit Iceberg verwenden	5131
Verwenden Sie einen Iceberg-Cluster mit Spark	5131
Verwenden Sie einen Iceberg-Cluster mit Trino	5136
Verwenden Sie einen Iceberg-Cluster mit Flink	5138
Verwenden Sie einen Iceberg-Cluster mit Hive	5143
Überlegungen und Einschränkungen	5146
Überlegungen zur Verwendung von Iceberg mit Spark	5146
Überlegungen zur Verwendung von Iceberg mit Trino	5147
Überlegungen zur Verwendung von Iceberg mit Flink	5147
Überlegungen zur Verwendung von Iceberg mit Hive	5147
Versionsverlauf von Iceberg	5148
Iceberg-Versionshinweise nach Version	5149
Jupyter Notebook	5151
EMR Studio	5151
EMR Notebook	5151
JupyterHub	5152
Erstellen Sie einen Cluster mit JupyterHub	5156
Überlegungen bei der Verwendung JupyterHub bei Amazon EMR	5158
Konfiguration JupyterHub	5159
Konfigurieren von Persistenz für Notebooks in Amazon S3	5160
Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen	5161
JupyterHub Konfiguration und Verwaltung	5162

Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren	5164
Installieren von zusätzlicher Kernen und Bibliotheken	5176
JupyterHub Versionsverlauf	5180
Livy	5217
Aktivieren HTTPS	5219
Livy-Versionsverlauf	5220
MXNet	5264
MXNetVerlauf der Veröffentlichung	5266
Oozie	5295
Verwenden von Oozie mit einer Remote-Datenbank in Amazon RDS	5298
Die Java-Version für Oozie konfigurieren	5300
Versionsverlauf von Oozie	5301
Versionshinweise zu Oozie nach Version	5358
Phoenix	5360
Einen Cluster mit Phoenix erstellen	5363
Anpassen von Phoenix-Konfigurationen	5364
Phoenix-Clients	5365
Versionsverlauf von Phoenix	5369
Pig	5443
Übermitteln von Pig-Aufträgen	5445
Arbeiten von Pig über die EMR Amazon-Konsole einreichen	5446
Reichen Sie Pig-Arbeiten ein, indem Sie AWS CLI	5447
Benutzerdefinierte Funktionen in Pig aufrufen	5449
Rufen Sie JAR Dateien von Pig aus auf	5449
Aufrufen von Python/Jython-Skripts in Pig	5449
Pig-Versionsverlauf	5451
Presto und Trino	5516
Presto mit dem AWS Glue-Datenkatalog verwenden	5519
AWS Glue Data Catalog als Metastore angeben	5520
IAM-Berechtigungen	4850
Überlegungen zur Verwendung von AWS Glue Data Catalog	5525
Verwenden von S3 Select-Pushdown	5526
Ist S3-Select-Pushdown das Richtige für meine Anwendung?	5526
Überlegungen und Einschränkungen	5526
Aktivieren von S3-Select-Pushdown für Presto oder Trino	5527
Hinzufügen von Datenbank-Konnektoren	5528

Verwenden vonSSL/TLSund LDAPS	5529
Authentifizierung verwenden LDAP	5530
Den strikten Presto-Modus aktivieren	5538
Überlegungen	5540
Umgang mit dem Verlust von Spot Instances in Presto	5541
Fehlertolerante Ausführung	5543
Konfiguration	5543
Austausch-Manager	5544
Überlegungen und Einschränkungen	5546
Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden	5546
Überlegungen zu Presto bei Amazon EMR	5547
In der Befehlszeile von Presto ausführbare Datei	5547
Nicht konfigurierbare Presto-Bereitstellungseigenschaften	5548
Installation von PrestoDB und Trino	5549
EMRFSund PrestoS3-Konfiguration FileSystem	5549
Standardeinstellung für Endbenutzer-Identitätswechsel	5550
Standardport für die Presto-Webschnittstelle	5551
Problem bei der Hive-Bucket-Ausführung in einigen Versionen	5551
Presto-Versionsverlauf	5551
Versionshinweise zu Trino (PrestoSQL) nach Version	5617
Spark	5620
Einen Spark-Cluster erstellen	5624
Führen Sie Spark-Anwendungen mit Docker auf Amazon EMR 6.x aus	5627
Überlegungen	5627
Ein Docker-Image erstellen	5628
Docker-Images von Amazon verwenden ECR	5629
Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL	5635
AWS Glue Data Catalog als Metastore angeben	5636
IAMBerechtigungen	4850
Überlegungen	4852
Konfigurieren von Spark	5640
Spark-Standardwerte	5642
Garbage Collection einrichten	5643
maximizeResourceAllocation	5644
Verhalten bei Außerbetriebnahme	5646
ThriftServer Spark-Umgebungsvariable	5650

Ändern von Spark-StandardEinstellungen	5651
Migration von Apache Log4j 1.x zu Log4j 2.x	5653
Die Leistung von Spark optimieren	5654
Adaptive Abfrageausführung	5654
Dynamische Partitionsbereinigung	5657
Abflachen skalarer Unterabfragen	5659
DISTINCTvorher INTERSECT	5660
Bloomfilter für Joins	5661
Optimierte Join-Neuanordnung	5661
Caching von Ergebnisfragmenten	5662
Caching von Ergebnisfragmenten in Spark aktivieren	5663
Überlegungen	5664
Verwenden Sie RAPIDS Accelerator	5666
Instance-Typen auswählen	5666
App-Konfigurationen einrichten	5666
Eine Bootstrap-Aktion hinzufügen	5674
Ihren Cluster starten	5675
Zugriff auf die Spark-Shell	5675
Verwenden Sie Amazon SageMaker Spark für maschinelles Lernen	5677
Eine Spark-Anwendung schreiben	5677
Scala	5677
Java	5678
Python	5679
Spark-Leistung mit S3 verbessern	5681
S3 Select verwenden	5681
EMRFSS3-optimierter Committer	5686
Verwenden Sie das S3-optimierte Commit-Protokoll EMRFS	5693
S3-Anforderungen wiederholen	5700
Einen Spark-Schritt hinzufügen	5703
Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen	5706
Anzeigen des Spark-Anwendungsverlaufs	5707
Greifen Sie auf das Spark-Web zu UIs	5707
Verwenden von Spark mit Amazon Kinesis Data Streams	5708
Spark auf Amazon Redshift verwenden	5708
Eine Spark-Anwendung starten	5709
Bei Amazon Redshift authentifizieren	5710

In Amazon Redshift schreiben und lesen	5712
Überlegungen	5714
Spark-Versionsverlauf	5715
Sqoop	5779
Sqoop-Versionen	5779
Überlegungen zu Sqoop auf Amazon EMR	5781
Verwenden von Sqoop mit Integration HCatalog	5781
Unterstützung für Sqoop JDBC und Datenbanken	5782
Schützen Ihres Passworts	5783
Versionsverlauf von Sqoop	5784
TensorFlow	5838
TensorFlow Builds nach EC2 Amazon-Instanztyp	5840
Sicherheit	5840
Verwenden TensorBoard	5840
TensorFlow Verlauf der Veröffentlichung	5841
Tez	5864
Erstellen eines Clusters mit Tez	5866
Konfigurieren von Tez	5867
Beispielkonfiguration	5867
Asynchrone Öffnung von Tez-Splits	5868
Tez-Webbenutzeroberfläche	5870
Timeline-Server	5871
Tez-Versionsverlauf	5871
Tez-Versionshinweise nach Version	5915
Zeppelin	5922
Überlegungen zur Verwendung von Zeppelin bei Amazon EMR	5924
Zeppelin-Versionshistorie	5925
ZooKeeper	5983
ZooKeeper Verlauf der Veröffentlichung	5985
Konnektoren und Dienstprogramme	6024
Tabellen in DynamoDB exportieren, importieren, abfragen und verknüpfen	6024
Eine Hive-Tabelle einrichten, um Hive-Befehle auszuführen	6026
Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten	6036
Optimierung der Leistung	6046
Kinesis	6050
Was kann ich mit der Integration von Amazon EMR und Amazon Kinesis machen?	6050

Analyse von Amazon-Kinesis-Streams mit Checkpointing	6050
Leistungsaspekte	6052
Planen Sie die Amazon Kinesis Kinesis-Analyse mit Amazon EMR	6052
Migration des Spark Kinesis-Connectors auf SDK 2.x für Amazon 7.0 EMR	6053
S3 DistCp (s3-dist-cp)	6062
DistCp S3-Optionen	6063
S3 DistCp als Schritt in einem Cluster hinzufügen	6071
Aufräumen nach fehlgeschlagenen DistCp S3-Jobs	6073
Befehle und Skripte auf einem Cluster ausführen	6075
Reichen Sie einen benutzerdefinierten JAR-Schritt ein, um ein Skript oder einen Befehl auszuführen	6076
Andere Verwendungsmöglichkeiten von <code>command-runner.jar</code>	6077
AWS-Glossar	6079
.....	6080

Über Amazon EMR Releases

Eine EMR Amazon-Version ist eine Reihe von Open-Source-Anwendungen aus dem Big-Data-Ökosystem. Jede Version umfasst verschiedene Big-Data-Anwendungen, -Komponenten und -Funktionen, die Sie bei der Erstellung eines Clusters für die EMR Installation und Konfiguration von Amazon auswählen. Anwendungen werden mithilfe eines Systems gepackt, das auf [Apache](#) basiert. Dabei handelt es sich um ein Open-Source-Projekt BigTop, das mit dem Hadoop-Ökosystem verknüpft ist. Dieses Handbuch enthält Informationen zu Anwendungen, die in EMR Amazon-Versionen enthalten sind.

Weitere Informationen zu den ersten Schritten und zur Arbeit mit Amazon EMR finden Sie im [Amazon EMR Management Guide](#).

Wenn Sie einen Cluster starten, können Sie aus mehreren Versionen von Amazon wählen EMR. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.2.0`.

Ab Amazon EMR 5.18.0 können Sie das Amazon EMR Artifact Repository verwenden, um Ihren Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die in bestimmten Amazon-Versionen verfügbar sind. EMR Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Um Updates zu erhalten, wenn eine neue EMR Amazon-Version verfügbar ist, abonnieren Sie den [RSSFeed für EMR Amazon-Versionshinweise](#).

Aktuelle Versionsdetails, einschließlich Anwendungsversionen, Versionshinweise, Komponenten und Konfigurationsklassifizierungen der Amazon EMR 7.x-, 6.x- und 5.x-Serien:

- [EMRAmazon-Version 7.2.0](#)
- [EMRAmazon-Version 6.15.0](#)
- [EMRAmazon-Version 5.36.2](#)

Note

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen

Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Versionshinweise zu den neuesten EMR Amazon-Veröffentlichungen und eine Historie aller Versionen:

- [Neuerungen?](#)
- [EMRAmazon-Archiv mit Versionshinweisen](#)

Eine umfassende Historie der Anwendungsversionen in jeder EMR Amazon-Version:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Einzelheiten zu jeder EMR Amazon-Version und gegebenenfalls zu den Unterschieden zwischen den Veröffentlichungsserien:

- [Release-Versionen von Amazon EMR 7.x](#)
- [Release-Versionen von Amazon EMR 6.x](#)
- [Amazon EMR 5.x-Release-Versionen](#)
- [Amazon EMR 4.x-Release-Versionen](#)
- [Amazon EMR 2.x- und AMI 3.x-Versionen](#)

EMRStandardunterstützung von Amazon

EMRAmazon-Releases verstehen

Die EMRAmazon-Version ist das Vehikel, das die erforderliche Software bereitstellt EC2EKS, um Ihre Open-Source-Anwendungen auf oder serverlosen Plattformen auszuführen. Eine EMR Amazon-Version besteht aus drei Teilen: Runtime Environment, Core Engines und Extras.

- **Laufzeitumgebung** — Die Laufzeitumgebung umfasst das Betriebssystem, das als Teil des Amazon Machine Image (AMI) oder eines anderen Container-Images enthalten ist, wenn Sie Ihren Cluster oder Container starten (z. B. Amazon Linux 2023). Es enthält auch die Sprachlaufzeiten wie Amazon Corretto sowie andere verwandte Tools wie JDK, die das Image installiert.
- **Kern-Engines** — Dazu gehören Open-Source-Kernsoftware wie Apache Hive und Open-Table-Formate wie Apache Iceberg.
- **Extras** — Dazu gehören praktische Bibliotheken und Python-Pakete wie `mariaadb-connector-java` sowie Open-Source-Software wie Apache Pig.

Die von Ihnen ausgewählte EMR Amazon-Release-Version bündelt all diese Komponenten in einer neuen Version von Amazon Linux (AL) -basierten Amazon Machine Images (AMI) oder Container-Images. Amazon EMR verwaltet dieselbe AL-Version für alle Knoten eines Clusters, EKS Containers oder einer serverlosen Anwendung. Amazon EMR ist bestrebt, die neueste Runtime Environment innerhalb von 90 Tagen nach ihrer Veröffentlichung durch Amazon Linux und neue Open-Source-Versionen von Core Engines innerhalb von 90 Tagen nach der Upstream-Version zu erhalten. Extras werden nach Bedarf veröffentlicht. Sie finden die Liste der Software in den Versionshinweisen für jede EMR Amazon-Version.

Versionierung der Versionen

Um Ihnen zu helfen, den Umfang der Änderungen in den einzelnen Versionen zu verstehen, verwendet Amazon semantische Versionierung. Die semantische Versionierung folgt einem Format von `<major>.<minor>.<patch>`. Betrachten wir als Beispiel die Version 7.0.0.

Die erste Ziffer bezeichnet die Hauptversionsversion oder Hauptversion. Hauptversionen beinhalten in der Regel wesentliche Änderungen, Verbesserungen und neue Funktionen an Runtime Environment oder Core Engine, die nicht abwärtskompatibel sind. Die 7.x-Hauptversion verwendet beispielsweise Amazon Linux 2023 mit Amazon Corretto 17 JDK als Standard. Daher enthält die Version mehrere wichtige Änderungen im Vergleich zur 6.x-Hauptversion, die auf Amazon Linux 2 und standardmäßig mit Amazon Corretto 8 JDK ausgeliefert wird.

Die zweite Ziffer steht für die Nebenversion oder Nebenversion. Nebenversionen sind nicht abwärtskompatible Versionen, die inkrementelle Änderungen, Verbesserungen und Funktionen an Core Engines und Extras sowie neue Funktionen enthalten. 6.15 verwendet beispielsweise Apache Spark 3.4.1 im Vergleich zu 6.11.1, das mit Apache Spark 3.3.2 ausgeliefert wird.

Die dritte Ziffer steht für eine Patch-Version oder Patch-Version. Bei Patch-Releases handelt es sich um abwärtskompatible Versionen, die Fixes und Patch-Updates für Core Engine und Extras,

aber keine neuen Funktionen oder OSS Versionsupgrades enthalten. Zum Beispiel enthalten 6.11.1 und 6.11.0 beide dieselbe Apache Spark 3.3.2-Version. OSS Ab 6.6.x bietet die neueste Patch-Version den neuesten verfügbaren Runtime Environment-Patch. Runtime Environment hat keinen Einfluss auf die Patch-Version von AmazonEMR. Zum Beispiel wurde 6.11.1 vom 19. Dezember 2023 bis 10. Januar 2024 mit der AL-Version 2.0.20240109.0 gestartet, ab dem 11. Januar 2024 jedoch mit der AL-Version 2.0.20240124.0. In einigen EMR EKS Amazon-Versionen können Sie den semantischen Versionen Tags oder Suffixe hinzufügen, um Ihnen alternative Funktionen zu bieten. Mit der `emr-6.15.0-latest` Version von Amazon EMR on EKS wird beispielsweise eine für Amazon Corretto 8 JDK aktivierte Release-Version gestartet, während die Version von Amazon EMR on EKS es Ihnen ermöglicht, eine für Amazon Corretto 17 JDK aktivierte `emr-6.15.0-java17-latest` Release-Version zu starten.

Support-Richtlinie

Einführung

Amazon EMR ist bestrebt, mindestens einmal alle 90 Tage neue Nebenversionen herauszugeben und Support für Nebenversionen für einen Zeitraum von 24 Monaten ab dem Veröffentlichungsdatum anzubieten. Dieser Support deckt Runtime Environment und Core Engines mit ihren zugehörigen Abhängigkeiten ab, nicht aber Extras oder die damit verbundenen Abhängigkeiten. Runtime Environment umfasst auch proprietäre Plattformkomponenten APIs, die zum Starten, Stoppen und Betrieb der Cluster und Anwendungen erforderlich sind. Dieser konsistente Release-Zeitplan gewährleistet einen vorhersehbaren Zyklus, sodass Sie leichter planen, testen und auf eine Version umsteigen können, die Support und Sicherheit bietet.

Was Sie vom Standard-Support erwarten können

Der Standardsupport bietet Lösungen für technische Supporttickets für Probleme, die bei Runtime Environment - und Core Engines-Komponenten im Rahmen der empfohlenen Konfigurationen auftreten. Alle Fixes unterliegen der Verfügbarkeit. Die empfohlene Konfiguration ist die Verwendung von Amazon EMR ohne Änderungen, Ergänzungen oder Änderungen an den Binärdateien und Konfigurationen, die in einer EMR Amazon-Version enthalten sind, mit Ausnahme der in der [EMR Amazon-Dokumentation](#) enthaltenen. Amazon EMR stellt innerhalb von 90 Tagen nach Überprüfung des Fixes Fixes für die neueste Patch-, Neben- oder Hauptversion der EMR Amazon-Version bereit. Amazon wendet EMR automatisch Fixes EMR an, wenn Sie einen neuen EC2 On-Cluster starten, einen neuen EMR EKS Amazon-On-Container starten oder einen neuen EMR Serverless-Job auslösen. Zusätzliche Komponenten werden der Einfachheit halber bereitgestellt, und Amazon EMR bietet keine Korrekturen im Zusammenhang mit Extras an.

Komponenten

Der Standardsupport deckt die Komponenten Runtime Environment und Core Engines ab, z. B. Betriebssystem, Sprachlaufzeiten und Open-Source-Kernsoftware wie Apache Hive und Apache Iceberg. Die vollständige Liste der unterstützten Komponenten für jede Version finden Sie in [Über Amazon EMR Releases](#) und in [Amazon Linux FAQs](#).

In der folgenden Liste wird der Support beschrieben, den wir für verschiedene Komponententypen unter Standard-Support anbieten:

- **Komponenten von Runtime Environment:** Die Komponenten von Runtime Environment werden im Rahmen von Tickets für den technischen Support behoben. Fixes werden in (a) kritische Fehler, (b) kritische Datenbeschädigung und (c) kritische Sicherheitsprobleme eingeteilt. Sofern berechtigt, EMR wird Amazon Fixes für Runtime Environment auf ältere Versionen zurückportieren. Aus Gründen der betrieblichen Kompatibilität mit den Open-Source-Core Engine-Komponenten müssen bestimmte Runtime Environment-Komponenten weiterhin in bestimmten Versionen verfügbar sein, um zu verhindern, dass Kundenanwendungen beschädigt werden. Für diese Komponenten EMR ist Amazon für die Verfügbarkeit von Fixes auf Upstream-Open-Source-Lösungen angewiesen. Wenn Fixes in Open Source verfügbar sind, stellen wir die neueste stabile Version innerhalb von 90 Tagen nach ihrer Überprüfung durch Amazon zur Verfügung EMR.
- **Core Engines-Komponenten:** Core Engines bietet die neuesten Versionen für viele Open-Source-Projekte, von denen jedes über Hunderte von transitiven Abhängigkeitsbibliotheken verfügt. Die Open-Source-Communities, die diese Projekte verwalten, versuchen zwar häufig, Probleme und bekannte allgemeine Sicherheitslücken und Sicherheitslücken (CVE) zu beheben, aber die neuesten Versionen enthalten möglicherweise immer noch bekannte Fehler und CVEs Amazon EMR ist bei der Verfügbarkeit von Fixes auf Upstream-Open-Source-Software angewiesen und wird die neueste stabile Version als Teil der Core Engine-Komponenten innerhalb von 90 Tagen nach deren Überprüfung durch Amazon EMR bereitstellen. In bestimmten Fällen EMR kann Amazon einen Fix für ein Problem CVE in einer der Core Engine-Komponenten bereitstellen, das vor dem Upstream-Open-Source-Prozess behoben werden muss. Amazon bietet Ihnen EMR außerdem technischen Support und Problembehebungen für Funktionen, die im Rahmen der empfohlenen Konfigurationen zusätzlich zu Open Source hinzugefügt werden. Wir führen keine Rückportierung von Fixes für Core Engine-Komponenten auf ältere Patch- oder Nebenversionen durch.
- **Extras-Komponenten:** Amazon unterstützt EMR keine Extras-Komponenten. Zusätzliche Komponenten sind Open-Source-Projekte, die der Einfachheit halber bereitgestellt werden, und Amazon bietet EMR keine Lösungen für Probleme, die mit ihnen auftreten. Alle Support-Anfragen

oder Problemlösungen können von der Open-Source-Community beantwortet werden, die diese Komponenten unterstützt.

Standard-Support-Lebenszyklus

Im Folgenden werden die Meilensteine im Standard-Support-Lebenszyklus beschrieben:

- **Standard-Support:** EMR Amazon-Versionen sind 24 Monate ab dem Veröffentlichungsdatum für Standard-Support qualifiziert. Sie können Tickets für den technischen Support erstellen und Updates für Probleme erwarten, die bei diesen Versionen auftreten.
- **Ende des Support:** Nach Ablauf des Standard-Supports treten EMR Amazon-Versionen für 12 Monate in die Phase des Endes des Support (EoS) über. Für eOS-Versionen besteht kein Anspruch auf technischen Support, und Sie können keine Tickets für Cluster, Container oder Jobs erstellen, die auf diesen Versionen ausgeführt werden. EoS-Versionen erhalten keine Fixes, Patches oder Updates. EoS-Versionen werden von der Konsole entfernt, sind aber weiterhin über API und verfügbar AWS CLI. Sie können jederzeit weiterhin Workloads auf EoS-Versionen ausführen. Wir empfehlen Ihnen dringend, auf die neueste EMR Amazon-Version zu migrieren, damit Sie weiterhin Sicherheitspatches erhalten, weiterhin Anspruch auf technischen Support haben und bei Bedarf Support-Tickets erstellen können.
- **Ende des Lebenszyklus:** Nach Ablauf der Laufzeit des Support (EoS) gelten Releases als End of Life (EoL). Obwohl Sie weiterhin EoL-Cluster ausführen können, behält sich Amazon das Recht vor, EoL-Releases aus Sicherheits API - und SDK Betriebsgründen nach und nach aus dem und zu entfernen. case-by-case Wir empfehlen dringend, auf die neueste Amazon-Version zu migrierenEMR, da EoL-Versionen SDK in Ausnahmefällen aus der API und entfernt werden können.

Bridge-Unterstützung

Amazon EMR kündigte diese neue Support-Richtlinie am 25. Juli 2024 an. Gemäß dieser Richtlinie gelten für Versionen von Amazon, die am oder vor dem 24. Juli 2022 EMR veröffentlicht wurden, nun das Ende des Support. Um Ihnen jedoch zusätzliche Zeit für die Planung und Migration zu neueren Versionen zu EMR geben, bietet Amazon Bridge Support, der dem Standard-Support entspricht, für diese älteren Versionen an, die innerhalb der zwei Jahre vor dieser Ankündigung veröffentlicht wurden.

Nach dem 25. Juli 2024 können Sie sich in den Versionshinweisen über den aktuellen Status und die Supportzeitpläne informieren.

Die folgende Tabelle zeigt den Support-Status für alle bestehenden EMR Amazon-Versionen zum Zeitpunkt der Ankündigung der Richtlinie am 25. Juli 2024:

Versionen und unterstützte Zeiträume

EMRAmazon-Veröffentlichungsversion	Erstes Veröffentlichungsdatum	Enddatum des Standard-Supports	Startdatum für das Ende des Supports	Startdatum am Ende der Laufzeit
7.2.0	25. Juli 2024	24. Juli 2026	25. Juli 2026	25. Juli 2027
7.1.0	23. April 2024	22. April 2026	23. April 2026	23. April 2027
7.0.0	19. Dezember 2023	18. Dezember 2025	19. Dezember 2025	19. Dezember 2026
5.36.x und 6.6.x — 6.15.x	9. Mai 2022 bis 13. November 2023	Bridge-Support bis 24. Januar 2026	25. Januar 2026	25. Januar 2026
<ul style="list-style-type: none"> 6.x-Serie : 6.5.0 und niedriger 5.x-Serie: 5.35.0 und niedriger Serien 4.x, 3.x und 2.x 	1. Januar 2013 bis 30. März 2022	Bridge-Support bis 24. Juli 2025	25. Juli 2025	25. Juli 2025

Überlegungen

Standard-Support ist für alle EMR Amazon-Bereitstellungsmodelle (EMRonEC2, Amazon EMR on EKS und EMR Serverless) in allen Regionen verfügbar, in denen Amazon verfügbar EMR ist, ohne zusätzliche Kosten. Cluster, die mit den empfohlenen Konfigurationen ausgeführt werden, kommen automatisch für den Support in Frage, wie in der Richtlinie beschrieben, sodass Sie keine zusätzlichen Maßnahmen ergreifen müssen, um den Support zu aktivieren.

- Der Standard-Support unterstützt nur Komponenten, die für EMR Amazon-Cluster erforderlich sind. Amazon EMR kann die Verfügbarkeit von Sicherheitspatches und Fixes nicht garantieren, falls die Open-Source-Komponenten der Core Engine den EoL-Upstream erreichen oder wenn Sicherheitsupdates für Abhängigkeiten nicht mehr verfügbar sind. Sie können sich zwar für die Installation von Extras entscheiden, Amazon unterstützt sie oder ihre Abhängigkeiten jedoch EMR nicht. Sie können beispielsweise Drittanbieteranwendungen in Ihrem eigenen System installieren, AMI um die Sicherheit Ihres Clusters zu erhöhen, zusätzliche Komponenten installieren oder Objekte mithilfe von Bootstrap-Aktionsskripten kopieren oder SSH in Ihren Cluster einbinden und die Standardpaketversionen aktualisieren. Amazon EMR unterstützt diese Komponenten nicht. Der Standard-Support deckt keine vom Kunden bereitgestellten Bootstrap-Aktionen, Pakete, Bibliotheken, Ihren benutzerdefinierten Code und bring-your-own benutzerdefinierte Anwendungen ab, die Sie Amazon so konfigurieren können, dass EMR sie bequem installiert werden.
- Ihre vorhandenen Cluster werden nicht beeinträchtigt, unabhängig davon, welche EMR Amazon-Version sie ausführen. Sie können bestehende Cluster weiterhin unterbrechungsfrei ausführen. Sie können auch weiterhin neue Cluster starten und Jobs für alle vorhandenen und neuen Versionen ausführen. Alle zum Zeitpunkt des Inkrafttretens der Richtlinie vorhandenen und neuen Versionen sind ab dem ersten Veröffentlichungsdatum der EMR Amazon-Release-Version 24 Monate lang durch den Standard-Support abgedeckt. Amazon bietet EMR Ihnen bei der ersten Ankündigung der Richtlinie Bridge-Support. Um ununterbrochenen Support zu erhalten, empfehlen wir, Ihre Anwendungen zu testen und umgehend ein Upgrade auf die neueste EMR Amazon-Version durchzuführen.
- Amazon EMR wird die Komponenten des Standard-Supports in bestehenden Versionen oder Clustern nicht ändern. Amazon EMR behält sich jedoch das Recht vor, das Upstream-Ende des Lebenszyklus auf einer bestimmten case-by-case Grundlage zu berücksichtigen und solche Komponenten in den neuen Versionen zu entfernen oder die bestehende Version in Ausnahmefällen in den Status End of Support (EoS) oder End of Life (EoL) zu versetzen. Wir werden Sie über die verfügbaren Kanäle über jede Entfernung informieren.
- Wenn eine neue Region eingeführt wird, Support Amazon nur EMR Amazon-Release-Versionen im Rahmen des Standard-Supports, die innerhalb von sechs Monaten vor dem Datum veröffentlicht wurden, an dem die neue Region allgemein verfügbar wird.
- Amazon aktualisiert Ihre vorhandenen Cluster EMR nicht automatisch auf die neuesten Versionen. Sie können sich jedoch dafür entscheiden, neue Cluster auf die neuesten Patch-Versionen zu aktualisieren, falls gewünscht.
- Die Kosten für Ihre nicht unterstützte EMR Nutzung von Amazon werden auf Ihre AWS Rechnung angerechnet. Auch wenn Sie Amazon auf eine EMR Weise nutzen, die nicht unterstützt wird, sind

die mit dieser Nutzung verbundenen Kosten dennoch Teil Ihres AWS Gesamtverbrauchs und werden in der Berechnung Ihrer Support-Gebühren berücksichtigt.

Für weitere Informationen wenden Sie sich an den [AWS Developer Support](#).

Release-Versionen von Amazon EMR 7.x

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon EMR 7.x-Release-Version verfügbar sind.

Wenn Sie einen Cluster starten, können Sie aus mehreren Versionen von Amazon wählen EMR. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. Beispiel, `emr-7.2.0`.

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon EMR 7.x-Version finden Sie unter [Anwendungsversionen in Amazon EMR 7.x-Versionen](#).

Themen

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [EMR Amazon-Version 7.2.0](#)
- [EMR Amazon-Version 7.1.0](#)
- [EMR Amazon-Version 7.0.0](#)

Anwendungsversionen in Amazon EMR 7.x-Versionen

Eine umfassende Tabelle mit den in jeder Amazon EMR 7.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon EMR 7.x-Versionen](#) in Ihrem Browser.

EMRAmazon-Version 7.2.0

7.2.0 unterstützter Lebenszyklus

In der folgenden Tabelle werden die unterstützten Lebenszyklusdaten für EMR Amazon-Version 7.2.0 beschrieben.

Unterstützungsphase	Datum
Erstes Veröffentlichungsdatum	25. Juli 2024
Standard-Support bis	24. Juli 2026
Ende des Supports	25. Juli 2026
Ende der Lebensdauer	25. Juli 2027

Eine Liste kritischer und häufig vorkommender Sicherheitslücken und Sicherheitslücken (CVEs), von denen EMR Cluster mit der empfohlenen Software und Konfiguration nicht betroffen sind, finden Sie unter [7.2.0 Bekannte häufige Sicherheitslücken und Sicherheitslücken für Core-Engines](#). Eine Liste der Probleme CVEs, die in Version 7.2.0 behoben wurden, finden Sie unter [7.2.0](#) behebt häufig auftretende Sicherheitslücken und Sicherheitslücken.

7.2.0 häufig auftretende Sicherheitslücken und Sicherheitslücken

Diese Seite listet alle gängigen Sicherheitslücken und Risiken (CVEs) im Zusammenhang mit Amazon EMR 7.2.0 auf.

7.2.0 bekannte allgemeine Sicherheitslücken und Sicherheitslücken für Core-Engines

In der folgenden Tabelle sind alle aufgeführt CVEs, die sich nicht auf EMR Cluster auswirken, die auf Amazon EMR 7.2.0 ausgeführt werden. Amazon EMR ist bei der Verfügbarkeit von Fixes auf Upstream-Open-Source-Software angewiesen und wird innerhalb von 90 Tagen, nachdem Amazon die Fixes EMR überprüft hat, die neueste stabile Version als Teil der Core Engine-Komponenten bereitstellen.

CVEID	Schweregrad	CVEEinzelheiten URL
CVE-2024-1597	Kritisch	https://nvd.nist.gov/vuln/detail/-2024-1597 CVE
CVE-2023-44981	Kritisch	https://nvd.nist.gov/vuln/detail/-2023-44981 CVE
CVE-2023-31047	Kritisch	https://nvd.nist.gov/vuln/detail/-2023-31047 CVE
CVE-2023-25668	Kritisch	https://nvd.nist.gov/vuln/detail/-2023-25668 CVE
CVE-2023-25664	Kritisch	https://nvd.nist.gov/vuln/detail/-2023-25664 CVE
CVE-2022 46337	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-46337 CVE
CVE-2022 42889	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-42889 CVE
CVE-2022 39135	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-39135 CVE
CVE-2022 - 37865	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-37865 CVE
CVE-2022 26612	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-26612 CVE
CVE-2022 25168	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-25168 CVE
CVE-2022—1471	Kritisch	https://nvd.nist.gov/vuln/detail/-2022-1471 CVE

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

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

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

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

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

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

CVEID	Schweregrad	CVEEinzelheiten URL
CVE-2023-25658	HIGH	https://nvd.nist.gov/vuln/detail/-2023-25658 CVE
CVE-2023-25194	HIGH	https://nvd.nist.gov/vuln/detail/-2023-25194 CVE
CVE-2023-24816	HIGH	https://nvd.nist.gov/vuln/detail/-2023-24816 CVE
CVE-2023-24580	HIGH	https://nvd.nist.gov/vuln/detail/-2023-24580 CVE
CVE-2023-23969	HIGH	https://nvd.nist.gov/vuln/detail/-2023-23969 CVE
CVE-2023-20883	HIGH	https://nvd.nist.gov/vuln/detail/-2023-20883 CVE
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/-2023-1436 CVE
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/-2023-1370 CVE
CVE-2022 - 46751	HIGH	https://nvd.nist.gov/vuln/detail/-2022-46751 CVE
CVE-2022 45693	HIGH	https://nvd.nist.gov/vuln/detail/-2022-45693 CVE
CVE-2022—45685	HIGH	https://nvd.nist.gov/vuln/detail/-2022-45685 CVE
CVE-2022 44729	HIGH	https://nvd.nist.gov/vuln/detail/-2022-44729 CVE
CVE-2022 42969	HIGH	https://nvd.nist.gov/vuln/detail/-2022-42969 CVE

CVEID	Schweregrad	CVEEinzelheiten URL
CVE-2022 42890	HIGH	https://nvd.nist.gov/vuln/detail/-2022-42890 CVE
CVE-20224-2004	HIGH	https://nvd.nist.gov/vuln/detail/-42-42004 CVE
CVE- 24.04.2003	HIGH	https://nvd.nist.gov/vuln/detail/-42-42003 CVE
CVE-2022 41704	HIGH	https://nvd.nist.gov/vuln/detail/-2022-41704 CVE
CVE-2022 40899	HIGH	https://nvd.nist.gov/vuln/detail/-2022-40899 CVE
CVE-2022—40664	HIGH	https://nvd.nist.gov/vuln/detail/-2022-40664 CVE
CVE-2022—40152	HIGH	https://nvd.nist.gov/vuln/detail/-2022-40152 CVE
CVE-2022—40150	HIGH	https://nvd.nist.gov/vuln/detail/-2022-40150 CVE
CVE-2022 40149	HIGH	https://nvd.nist.gov/vuln/detail/-2022-40149 CVE
CVE-2022—40146	HIGH	https://nvd.nist.gov/vuln/detail/-2022-40146 CVE
CVE-2022 37866	HIGH	https://nvd.nist.gov/vuln/detail/-2022-37866 CVE
CVE-2022 36364	HIGH	https://nvd.nist.gov/vuln/detail/-2022-36364 CVE
CVE-2022 3510	HIGH	https://nvd.nist.gov/vuln/detail/-2022-3510 CVE

CVEID	Schweregrad	CVEEinzelheiten URL
CVE-2022 3509	HIGH	https://nvd.nist.gov/vuln/detail/-2022-3509 CVE
CVE-2022 - 34169	HIGH	https://nvd.nist.gov/vuln/detail/-2022-34169 CVE
CVE-2022 32532	HIGH	https://nvd.nist.gov/vuln/detail/-2022-32532 CVE
CVE-2022 3171	HIGH	https://nvd.nist.gov/vuln/detail/-2022-3171 CVE
CVE-2022 25647	HIGH	https://nvd.nist.gov/vuln/detail/-2022-25647 CVE
CVE-2022 24758	HIGH	https://nvd.nist.gov/vuln/detail/-2022-24758 CVE
CVE-2021-41303	HIGH	https://nvd.nist.gov/vuln/detail/-2021-41303 CVE
CVE-2021-37137	HIGH	https://nvd.nist.gov/vuln/detail/-2021-37137 CVE
CVE-2021-37136	HIGH	https://nvd.nist.gov/vuln/detail/-2021-37136 CVE
CVE-2021-34538	HIGH	https://nvd.nist.gov/vuln/detail/-2021-34538 CVE
CVE-2021-33813	HIGH	https://nvd.nist.gov/vuln/detail/-2021-33813 CVE
CVE-2021-33036	HIGH	https://nvd.nist.gov/vuln/detail/-2021-33036 CVE
CVE-2021-31684	HIGH	https://nvd.nist.gov/vuln/detail/-2021-31684 CVE

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

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

CVEID	Schweregrad	CVEEinzelheiten URL
CVE-2020-17533	HIGH	https://nvd.nist.gov/vuln/detail/-2020-17533 CVE
CVE-2020-17523	HIGH	https://nvd.nist.gov/vuln/detail/-2020-17523 CVE
CVE-2020-14195	HIGH	https://nvd.nist.gov/vuln/detail/-2020-14195 CVE
CVE-2020-14062	HIGH	https://nvd.nist.gov/vuln/detail/-2020-14062 CVE
CVE-2020-14061	HIGH	https://nvd.nist.gov/vuln/detail/-2020-14061 CVE
CVE-2020-14060	HIGH	https://nvd.nist.gov/vuln/detail/-2020-14060 CVE
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/-2020-13949 CVE
CVE-2020-13936	HIGH	https://nvd.nist.gov/vuln/detail/-2020-13936 CVE
CVE-2020-1988	HIGH	https://nvd.nist.gov/vuln/detail/-2020-11988 CVE
CVE-2020-11620	HIGH	https://nvd.nist.gov/vuln/detail/-2020-11620 CVE
CVE-2020-11619	HIGH	https://nvd.nist.gov/vuln/detail/-2020-11619 CVE
CVE-2020-11113	HIGH	https://nvd.nist.gov/vuln/detail/-2020-11113 CVE
CVE-2020-11112	HIGH	https://nvd.nist.gov/vuln/detail/-2020-11112 CVE

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

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

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

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

7.2.0 hat häufig auftretende Sicherheitslücken und Sicherheitslücken für Core-Engines behoben

In der folgenden Tabelle sind alle Probleme aufgeführt, die CVEs in Amazon EMR 7.2.0 behoben wurden.

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

Komponenten der Version 7.2.0

Amazon EMR 7.2.0 unterstützt die folgenden Anwendungen:

Kern-Engines:

AmazonCloudWatchAgent, Delta, Flink, Hadoop,, Hive HBaseHCatalog, Hudi, Livy, Iceberg, Phoenix, Spark, Tez, JupyterEnterpriseGateway Trino, Tierpfleger

Zusätze:

Hue JupyterHub, OzieMXNet, Schwein, Presto, Sqoop, Zeppelin TensorFlow

[Weitere Informationen zu den unterstützten Paketen finden Sie unter Details zu den Komponenten der Version 7.2.0.](#) Weitere Informationen zu Core-Engines und Extras finden Sie unter [Amazon EMR Standard Support](#).

Einzelheiten zu den Komponenten der Version 7.2.0

In der folgenden Tabelle finden Sie weitere Informationen zu den Extras-Paketen in Amazon EMR 7.2.0.

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
AmazonCloudWatchAgent	Core	hadoop-hdfs-journalnode	Extras	3.3.6. amzn.4-1. amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
AmazonCloudWatchAgent	Core	Ranger-Kms	Komparsen	2.0.0-1.amzn2023
AmazonCloudWatchAgent	Core	s3-dist-cp	Komparsen	2.32.0-1.amzn2023
AmazonCloudWatchAgent	Core	Zoowärter	Extras	3.9.1.amzn.1—1.amzn2023
AmazonCloudWatchAgent	Core	zookeeper-server	Komparsen	3.9.1.amzn.1—1.amzn2023
Flink	Core	hadoop-client	Komparsen	3.3.6.amzn.4-1.amzn2023
Flink	Core	Hadoop-HDFS	Komparsen	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs-datanode	Komparsen	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs-journalnode	Komparsen	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-hdfs-namenode	Komparsen	3.3.6.amzn.4-1.amzn2023
Flink	Core	hadoop-httfs	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Core	Hadoop-km	Komparsen	3.3.6.amzn.4-1.amzn2023
Flink	Core	Hadoop-Ma preduce	Komparsen	3.3.6.amzn.4-1.amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
Flink	Core	hadoop-mapreduce-historyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Flink	Core	Hadoop-Garn	Komparsen	3.3.6. amzn.4-1. amzn2023
Flink	Core	hadoop-yarn-nodemanager	Komparsen	3.3.6. amzn.4-1. amzn2023
Flink	Core	hadoop-yarn-proxyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Flink	Core	hadoop-yarn-resourcemanager	Komparsen	3.3.6. amzn.4-1. amzn2023
Flink	Core	hadoop-yarn-timelineserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Flink	Core	Hudi	Komparsen	0.14.1.amzn.1-1.amzn2023
Flink	Core	Ranger-Kms	Komparsen	2.0.0-1.amzn2023
Flink	Core	Hadoop	Extras	3.3.6. amzn.4-1. amzn2023
HBase	Core	s3-dist-cp	Komparsen	2.32.0-1. amzn2023
HCatalog	Core	hadoop-yarn-proxyserver	Extras	3.3.6. amzn.4-1. amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
HCatalog	Core	mariadb-connector-java	Komparsen	2.7.2-1
Hive	Core	hadoop-ftpfs	Extras	3.3.6. amzn.4-1. amzn2023
Hive	Core	hadoop-yarn-proxyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Hive	Core	hive-hbase	Komparsen	3.1.3. amzn.11-1. amzn2023
Hive	Core	mariadb-connector-java	Komparsen	2.7.2-1
Hive	Core	s3-dist-cp	Komparsen	2.32.0-1. amzn2023
Phoenix	Core	s3-dist-cp	Extras	2.32.0-1. amzn2023
Spark	Core	Funken-Datenkern	Komparsen	3.5.1. amzn.0-1. amzn2023
Tez	Core	Hadoop-Ma preduce	Komparsen	3.3.6. amzn.4-1. amzn2023
Tez	Core	hadoop-mapreduce-historyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Tez	Core	hadoop-yarn-proxyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	bigtop-Tools	Komparsen	1.2.0-1. amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
Trino	Core	hadoop-client	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	Hadoop-HDFS	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-hdfs-datanode	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-hdfs-journalnode	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-hdfs-namenode	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	Hadoop-km	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	Hadoop-Ma preduce	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-ma preduce-historyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	Hadoop-Garn	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-yarn-nodemanager	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-yarn-proxyserver	Komparsen	3.3.6. amzn.4-1. amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
Trino	Core	hadoop-yarn-resource-manager	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-yarn-timelineserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Trino	Core	Hive	Komparsen	3.1.3. amzn.11-1. amzn2023
Trino	Core	hive-hcatalog-server	Komparsen	3.1.3. amzn.11-1. amzn2023
Trino	Core	mariadb-connector-java	Komparsen	2.7.2-1
Trino	Core	Ranger-Kms	Komparsen	2.0.0-1. amzn2023
Trino	Core	Zoowärter	Extras	3.9.1. amzn.1—1. amzn2023
Trino	Core	zookeeper-server	Komparsen	3.9.1. amzn.1—1. amzn2023
Trino	Core	bigtop-groovig	Komparsen	2.5.4-1. amzn2023
Trino	Core	bigtop-jsvc	Komparsen	1.2.4-1. amzn2023
Trino	Core	Hadoop	Extras	3.3.6. amzn.4-1. amzn2023
Trino	Core	hadoop-lzo	Komparsen	0.4.19-1. amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
Trino	Core	hive-h-Katalog	Extras	3.1.3. amzn.11-1 .amzn2023
Trino	Core	hive-jdbc	Komparsen	3.1.3. amzn.11-1 .amzn2023
Zookeeper	Core	emrfs	Komparsen	2.63.0-1. amzn2023
Zookeeper	Core	hadoop-client	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	Hadoop-HDFS	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-hdfs- datanode	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-hdfs- journalnode	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-hdfs- namenode	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-httfs	Extras	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	Hadoop-km	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	Hadoop-Ma preduce	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-ma preduce-h istoryserver	Komparsen	3.3.6. amzn.4-1. amzn2023

Anwendung	Art der Anwendung	Paket	Pakettyp	Version Package
Zookeeper	Core	Hadoop-Garn	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-yarn-nodemanager	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-yarn-proxyserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-yarn-resourcemanager	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	hadoop-yarn-timelineserver	Komparsen	3.3.6. amzn.4-1. amzn2023
Zookeeper	Core	Ranger-Kms	Komparsen	2.0.0-1. amzn2023

7.2.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK für 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-amzn-0	1.4.3-amzn-0	1.4.2-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

	emr-7.2.0	emr-7.1.0	emr-7.0.0
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
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

Versionshinweise zu 7.2.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 7.2.0. Die Änderungen beziehen sich auf 7.2.0.

Neue Features

- Anwendungs-Upgrades – Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.
- Amazon EMR fügt Unterstützung hinzu, sodass Sie andere Anwendungen wie HBase Flink und Hive mit der Amazon S3 Express One Zone-Speicherklasse verwenden können.
- Diese Version bietet die Möglichkeit, wiederhergestellte Objekte zu lesen, sodass Sie Glacier-Objekte mit dem Protokoll von einem S3-Standort aus lesen können. S3A Diese Funktion funktioniert mit Spark, Flink und Hive.

- Node-Labels mit verwalteter Skalierung — Verwenden Sie Node-Labels für Ihre Instances, die auf dem Markttyp oder dem Knotentyp basieren, um die automatische Skalierung von Amazon zu verbessern. Weitere Informationen finden Sie unter [Managed Scaling in Amazon verwenden](#) [EMR](#).

Bekannte Probleme

- Python 3.11 wird von EMR Studio nicht unterstützt.

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt ein Deadlock-Problem, das bei internen Cleanup-Vorgängen auftreten kann. Dieser Vorgang verwaltet den Lebenszyklus der einzelnen Schritte, während sie auf dem EMR Cluster abgeschlossen werden. Dieses Problem betrifft kritische EMR Amazon-Abläufe wie den schrittweisen Betrieb und die Skalierung.
- Diese Version behebt ein Problem, bei dem benutzerdefinierte Cluster mit benutzerdefinierten Clustern, die bereits über bestimmte Protokolldateien verfügen, dazu führen können, dass der Amazon EMR Log Management Daemon fehlschlägt.
- Amazon EMR 7.2.0 aktualisiert den EMR Amazon-Daemon, der für die Clusterverwaltung und die Überwachungsaktivitäten verantwortlich ist, von AWS SDK Version 1 auf Version 2.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.5.2 240708.1	6.1.96-102.177.amzn2023	8. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

7.2.0 Standard-Java-Versionen

Amazon EMR Version 7.2 und höher wird standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen ausgeliefert, die Corretto 17 (JDK17) unterstützen, mit Ausnahme von Apache Livy.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 7.2.0. Wenn Sie die Standardeinstellung JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
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

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Spark	17, 11, 8
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

7.2.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
adot-java-agent	1.31.0	Ein Java-Agent, der Metriken von Anwendungs-Daemons sammelt.
delta	3.1.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-amazon-cloudwatch-agent	1,300032,2-amzn-0	Eine Anwendung, die interne Metriken auf Systemebene und benutzerdefinierte Anwendungsmetriken von EC2 Amazon-Instances sammelt.
emr-ddb	5.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.15.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.16.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.32.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.11.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emr-wal-cli	1.2.2	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.63.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.18.1-amzn-1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.18.1-amzn-1	Verwaltung von Ressourcen auf Knoten für Apache Flink. EMR JobManager
hadoop-client	3.3.6-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.6-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.6-amzn-4	HDFSBefehlszeilen-Client und Bibliothek
hadoop-hdfs-namenode	3.3.6-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.6-amzn-4	HTTPEndpunkt für Operationen. HDFS

Komponente	Version	Beschreibung
hadoop-kms-server	3.3.6-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.3.6-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer Anwendung. MapReduce
hadoop-yarn-nodemanager	3.3.6-amzn-4	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.3.6-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.6-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für Bewerbungen. YARN
hbase-hmaster	2.4.17-amzn-6	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-6	Service für die Versorgung einer oder mehrerer Regionen. HBase
hbase-client	2.4.17-amzn-6	HBase-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hbase-rest-server	2.4.17-amzn-6	Dienst, der einen Endpunkt für bereitstellt. RESTful HTTP HBase
hbase-thrift-server	2.4.17-amzn-6	Dienst, der einen Thrift-Endpunkt für bietet. HBase
hbase-operator-tools	2.4.17-amzn-6	Reparatur-Tool für Apache-Cluster. HBase
hcatalog-client	3.1.3-amzn-11	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-11	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-11	HTTP-Endpunkt, der eine Schnittstelle zu bereitstellt. REST HCatalog
hive-client	3.1.3-amzn-11	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-11	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-11	Dienst für den Zugriff auf den Hive Metastore, ein semantisches Repository, das Metadaten für Hadoop-Operationen speichert. SQL
hive-server2	3.1.3-amzn-11	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.14.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0,14.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0,14.1-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0,14.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1,5.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.8.0-inkubieren	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	6.0.0	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0,285-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0,285-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0,285-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
trino-coordinator	436-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	436-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	436-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.3.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.5.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.5.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.5.1-amzn-0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	3.5.1-amzn-0	Apache Spark-Bibliotheken, die von Slaves benötigt werden. YARN
spark-rapids	24.02.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-9	Die TEZ-Anwendung und die Bibliotheken. YARN
tez-on-worker	0.10.2-amzn-9	Die TEZ-Anwendung und die Bibliotheken für Worker-Knoten YARN.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.9.1-amzn-1	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.9.1-amzn-1	ZooKeeper Befehlszeilen-Client.

7.2.0-Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-7.2.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> von Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändern Sie die Werte in der HDFS Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der <code>lakeformation.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratiert e Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-logback	Ändern Sie die Werte in ZooKeeper der Datei logback.xml.	Restarts Zookeeper server.
emr-metrics	Ändern Sie die emr-Metrikeinstellungen für diesen Knoten.	Restarts the CloudWatchAgent service.

7.2.0 Änderungsprotokoll

Änderungsprotokoll für die Version 7.2.0 und die Versionshinweise

Datum	Ereignis	Beschreibung
2024-07-25	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 7.2.0

Datum	Ereignis	Beschreibung
16.07.2024-07	Erstversion	Amazon EMR 7.2.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 7.1.0

7.1.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK für 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-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0

	emr-7.2.0	emr-7.1.0	emr-7.0.0
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

Versionshinweise zu 7.1.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 7.1.0. Die Änderungen beziehen sich auf 7.1.0.

Neue Features

- Anwendungs-Upgrades – Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1.
- [Austausch fehlerhafter Knoten](#) — Bei Amazon EMR 7.1.0 und höher ist der Austausch fehlerhafter Knoten standardmäßig aktiviert, sodass Amazon Ihre EMR fehlerhaften Knoten ordnungsgemäß ersetzt. Um zu vermeiden, dass Ihre bestehenden Workflows auf EMR Amazon-Versionen 7.0.0 und niedriger beeinträchtigt werden, ist der Austausch fehlerhafter Knoten deaktiviert, wenn Sie den Kündigungsschutz in Ihrem Cluster aktiviert haben.
- [CloudWatch Agent](#) — Konfigurieren Sie den CloudWatch Agenten so, dass er zusätzliche Systemmetriken verwendet, Anwendungsmetriken hinzufügt und das Metrikziel mit der EMR Amazon-Konfiguration ändertAPI.

Bekannte Probleme

- Python 3.11 wird von EMR Studio nicht unterstützt.

Änderungen, Verbesserungen und behobene Probleme

- Während Amazon EMR 7.1.0 Python 3.9 standardmäßig unterstützt, unterstützen Livy 0.8.0 und Spark in Amazon EMR 7.1.0 Python 3.11.
- Diese Version behebt das Problem, dass bei Verwendung PySpark mit Python Version 3.11 jede Zeile einzeln ausgeführt werden muss.
- Zeppelin-Upgrade — Amazon EMR 7.1.0 beinhaltet ein Upgrade von Zeppelin auf Java v2. AWS SDK Dieses Upgrade ermöglicht es einem Zeppelin S3-Notebook, einen Anbieter von kundenspezifischen Verschlüsselungsmaterialien zu akzeptieren. Das AWS SDK für Java v2 entfernt die EncryptionMaterialsProvider Schnittstelle. Wenn Sie ein Upgrade auf Amazon EMR 7.1.0 durchführen, müssen Sie die [Keyring-Schnittstelle](#) implementieren, wenn Sie eine benutzerdefinierte Verschlüsselung verwenden möchten. [Ein Beispiel für die Implementierung der Keyring-Schnittstelle finden Sie unter .java. KmsKeyring](#)
- Wenn Sie ein Upgrade auf EMR Amazon-Version 7.1.0 durchführen, ändern Sie Ihren benutzerdefinierten Schlüsselanbieter für die lokale Festplattenverschlüsselung, sodass Schlüssel mit dem AES Algorithmus von AES/GCM/NoPadding generiert werden. Wenn Sie den Algorithmus nicht aktualisieren, schlägt die Clustererstellung möglicherweise mit der folgenden Fehlermeldung fehl. Weitere Informationen `Local disk encryption failed on master instance (i-123456789) due to internal error.` zum Erstellen eines benutzerdefinierten Schlüsselanbieters finden Sie unter [Erstellen eines benutzerdefinierten Schlüsselanbieters](#).
- Amazon EMR 7.1.0 verbessert die Ausfallsicherheit eines Knotens bei geringem Festplattenspeicher, indem die Logik zur Protokollkürzung für Dateien mit offenen Datei-Handles verbessert wird.
- Diese Version verbessert die Kodierungs- und Dekodierungslogik, um das Risiko einer Datenbeschädigung und eines Knotenausfalls zu minimieren, EMR da Amazon-Daemons beim Neustart eines Nodes Dateien lesen und schreiben.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.5.240708.1	6.1.96-102.177.amzn2023	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240219.0	6.1.77-99.164.amzn2023	8. Mai 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

7.1.0 Standard-Java-Versionen

Amazon EMR Version 7.1 und höher wird standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen ausgeliefert, die Corretto 17 (JDK17) unterstützen, mit Ausnahme von Apache Livy.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 7.1.0. Wenn Sie die Standardeinstellung JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	17, 11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
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
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

7.1.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
adot-java-agent	1.31.0	Ein Java-Agent, der Metriken von Anwendungs-Daemons sammelt.
delta	3.0.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-amazon-cloudwatch-agent	1,300032,2-amzn-0	Eine Anwendung, die interne Metriken auf Systemebene und benutzerdefinierte Anwendungsmetriken von EC2 Amazon-Instances sammelt.
emr-ddb	5.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.10.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.14.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.31.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.10.0	EMRS3Select-Anschluss
emr-wal-cli	1.2.2	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.62.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.18.1-amzn-0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.18.1-amzn-0	Verwaltung von Ressourcen auf Knoten für Apache Flink. EMR JobManager
hadoop-client	3.3.6-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.6-amzn-3	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.6-amzn-3	HDFSBefehlszeilen-Client und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.3.6-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-3	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.6-amzn-3	HTTPEndpunkt für Operationen. HDFS
hadoop-kms-server	3.3.6-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API
hadoop-mapred	3.3.6-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer Anwendung. MapReduce
hadoop-yarn-nodemanager	3.3.6-amzn-3	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.6-amzn-3	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.6-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für Bewerbungen. YARN

Komponente	Version	Beschreibung
hbase-hmaster	2.4.17-amzn-5	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-5	Service für die Versorgung einer oder mehrerer Regionen. HBase
hbase-client	2.4.17-amzn-5	HBase Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-5	Dienst, der einen Endpunkt für bereitstellt. RESTful HTTP HBase
hbase-thrift-server	2.4.17-amzn-5	Dienst, der einen Thrift-Endpunkt für bietet. HBase
hbase-operator-tools	2.4.17-amzn-5	Reparatur-Tool für Apache-Cluster. HBase
hcatalog-client	3.1.3-amzn-10	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-10	Bereitstellung von Diensten HCatalog, eine Tabellen- und Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-10	HTTP-Endpunkt, der eine Schnittstelle zu bereitstellt. REST HCatalog
hive-client	3.1.3-amzn-10	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-10	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-10	Dienst für den Zugriff auf den Hive Metastore, ein semantisches Repository, das Metadaten für Hadoop-Operationen speichert. SQL
hive-server2	3.1.3-amzn-10	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.14.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.14.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.14.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.14.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.4.3-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.8.0-inkubieren	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0,284-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0,284-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0,284-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	435-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	435-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	435-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.3.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.5.0-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.5.0-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.5.0-amzn-1	In-Memory-Ausführung engine für YARN
spark-yarn-slave	3.5.0-amzn-1	Apache Spark-Bibliotheken, die von Slaves benötigt werden. YARN
spark-rapids	23.10.0-amzn-1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-8	Die TEZ-Anwendung und die Bibliotheken. YARN
tez-on-worker	0.10.2-amzn-8	Die TEZ-Anwendung und die Bibliotheken für Worker-Knoten YARN.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.9.1-amzn-0	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.9.1-amzn-0	ZooKeeper Befehlszeilen-Cli-ent.

7.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

EMR-7.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.
emrfs-site	EMRFSEinstellungen ändern.

Klassifizierungen	Beschreibung
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.

Klassifizierungen	Beschreibung
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-env	Ändern Sie die Werte in der HDFS Umgebung.
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. hive-exec-log
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.

Klassifizierungen	Beschreibung
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-logback	Ändern Sie die Werte in ZooKeeper der Datei logback.xml.
emr-metrics	Ändern Sie die emr-Metrikestellungen für diesen Knoten.

7.1.0 Änderungsprotokoll

Änderungsprotokoll für Version 7.1.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2024-05-10	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 7.1.0

Datum	Ereignis	Beschreibung
2024-05-08	Die Bereitstellung ist abgeschlossen	Amazon EMR 7.1.0 vollständig in allen unterstützten Regionen bereitgestellt
2024-04-23	Erstversion	Amazon EMR 7.1.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 7.0.0

7.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK für 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-amzn-0	1.4.3-amzn-0	1.4.2-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

7.0.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 7.0.0. Änderungen beziehen sich auf Version 6.15.0.

Neue Features

- Anwendungs-Upgrades – Amazon EMR 7.0.0 application upgrades include Python 3.9, Spark 3.5, Flink 1.18, and Delta 3.0. Diese Version bietet auch Unterstützung für die Anwendung [CloudWatch Amazon-Vertreter](#) und entfernt die Unterstützung für Ganglia.
- Amazon Corretto 17 — EMR Amazon-Versionen 7.0 und höher werden standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen ausgeliefert, die Corretto 17 (JDK17) unterstützen, mit Ausnahme von Apache Livy. Weitere Informationen zu den unterstützten JDK Versionen für Anwendungen in dieser Version finden Sie unter [7.0.0 Standard-Java-Versionen](#)
- Amazon Linux 2023 — Mit der Version 7.0 laufen EMR Amazon-Cluster jetzt standardmäßig auf AL2 023. Informationen dazu, wie sich dies auf die AMI Standardversion auswirkt, finden Sie unter

[Überlegungen zu Software-Updates](#) im Amazon EMR Management Guide. Beachten Sie auch, dass AL2 023 Python 2.7 entfernt hat, sodass alle Komponenten, die Python benötigen, jetzt mit Python 3 geschrieben werden sollten.

- S3 auf Outposts mit **s3a** — Amazon unterstützt EMR jetzt Amazon S3 auf Outposts-Buckets mit dem s3a Dateisystem. Weitere Informationen zu S3 in Outposts finden Sie unter [Was ist S3 in Outposts?](#) im Benutzerhandbuch für Amazon Simple Storage Service.

Bekannte Probleme

- Beachten Sie, dass Sie nicht mehr als eine Zeile gleichzeitig ausführen können, wenn Sie Python-Versionen 3.10 oder höher verwenden PySpark. Sie müssen jede Zeile einzeln ausführen.

Änderungen, Verbesserungen und behobene Probleme

- Um den Status aller Instances in einem Cluster aufrechtzuerhalten, EMR integriert Amazon ApacheYARN, Apache HDFS und Kerberos. Mit 7.0 haben wir diese Integrationen verbessert, um den Status von Instances, die aufgrund von Skalierung oder anderen Vorgängen beendet wurden, zuverlässig zu entfernen. Dies ist besonders wichtig für langlebige Cluster, bei denen die verwaltete Skalierung aktiviert ist, da sie häufig skalieren und im Laufe ihrer Lebensdauer Tausende beendete Instances ansammeln.
- In dieser Version wurde die Kerberos-Konfiguration dahingehend verbessert, dass nur noch basierte Verschlüsselungen unterstützt werden. AES Kerberos KDC mit nicht AES basierten Verschlüsselungen werden in EMR Clustern, die auf EMR Amazon-Versionen 7.0.0 und höher ausgeführt werden, nicht mehr unterstützt. AESbasierte Verschlüsselungen bieten die stärkste Sicherheit für Ihre Cluster.
- Im Rahmen der AWS SDK 2.x-Migration enthält Amazon EMR 7.0 aus Kompatibilitätsgründen ein Update für den Spark Kinesis-Konnektor. Dieses Update ist in der Community-Version von Apache Spark nicht verfügbar. Wenn Sie den Spark Kinesis-Connector von einer EMR Amazon-Version unter 7.0 verwenden, müssen Sie Ihre Anwendungscodes für die Ausführung auf SDK 2.x migrieren, bevor Sie Ihre Workloads auf Amazon 7.0 migrieren können. EMR Weitere Informationen finden Sie unter [Migration des Spark Kinesis-Connectors auf SDK 2.x für Amazon 7.0 EMR](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Arbeit-Arbeit-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.5.2 240708.1	6.1.96-102.177.amzn2023	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240304,0	6.1.79-99.164.amzn2023	12. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240219.0	6.1.77-99.164.amzn2023	1. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240205,0	6.1.75-99.163.amzn2023	19. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240122,1	6.1.72-96.amzn2023	5. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240108,0	6.1.72-96.amzn2023	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.2 231211,4	6.1.66-91.160.amzn2023	19. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

7.0.0 Standard-Java-Versionen

Amazon EMR Version 7.0 und höher wird standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen ausgeliefert, die Corretto 17 (JDK17) unterstützen, mit Ausnahme von Apache Livy.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 7.0.0. Wenn Sie die Standardeinstellung JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

7.0.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
adot-java-agent	1.31.0	Ein Java-Agent, der Metriken von Anwendungs-Daemons sammelt.
delta	3.0.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-amazon-cloudwatch-agent	1.300031.1-amzn-0	Eine Anwendung, die interne Metriken auf Systemebene und benutzerdefinierte Anwendungsmetriken von EC2 Amazon-Instances sammelt.
emr-ddb	5.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.9.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.13.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.30.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.9.0	EMRS3Select-Anschluss
emr-wal-cli	1.2.1	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.61.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.18.0-amzn-0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.18.0-amzn-0	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
hadoop-client	3.3.6-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.6-amzn-2	HDFSDienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.6-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.6-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-2	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.6-amzn-2	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.3.6-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.3.6-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.6-amzn-2	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.6-amzn-2	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.6-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.17-amzn-4	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-4	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.17-amzn-4	HBase Befehlszeilenclient.
hbase-rest-server	2.4.17-amzn-4	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.17-amzn-4	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-4	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-9	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-9	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	3.1.3-amzn-9	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-9	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.14.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.14.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.14.0-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.14.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.4.2-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.283-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.283-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.283-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	426-amzn-1	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	426-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
trino-client	426-amzn-1	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.1.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.5.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.5.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.5.0-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.5.0-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.10.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-7	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-7	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.58	Apache-ServerHTTP.

Komponente	Version	Beschreibung
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

7.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-7.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> YARN von Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase

Klassifizierungen	Beschreibung
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-env	Ändern Sie die Werte in der HDFS Umgebung.
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

7.0.0 Änderungsprotokoll

Änderungsprotokoll für Version 7.0.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-12-29	Die Bereitstellung ist abgeschlossen	Amazon EMR 7.0.0 wurde vollständig in allen unterstützten Regionen bereitgestellt
2023-12-28	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 7.0.0
2023-12-19	Erstversion	Amazon EMR 7.0.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

Release-Versionen von Amazon EMR 6.x

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon EMR 6.x-Release-Version verfügbar sind.

Wenn Sie einen Cluster starten, können Sie aus mehreren Versionen von Amazon wählen EMR. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.2.0`.

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon EMR 6.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#).

Themen

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)

- [EMRAmazon-Version 6.15.0](#)
- [EMRAmazon-Version 6.14.0](#)
- [EMRAmazon-Version 6.13.0](#)
- [EMRAmazon-Version 6.12.0](#)
- [EMRAmazon-Version 6.11.1](#)
- [EMRAmazon-Version 6.11.0](#)
- [EMRAmazon-Version 6.10.1](#)
- [EMRAmazon-Version 6.10.0](#)
- [EMRAmazon-Version 6.9.1](#)
- [EMRAmazon-Version 6.9.0](#)
- [EMRAmazon-Version 6.8.1](#)
- [EMRAmazon-Version 6.8.0](#)
- [EMRAmazon-Version 6.7.0](#)
- [EMRAmazon-Version 6.6.0](#)
- [EMRAmazon-Version 6.5.0](#)
- [EMRAmazon-Version 6.4.0](#)
- [EMRAmazon-Version 6.3.1](#)
- [EMRAmazon-Version 6.3.0](#)
- [EMRAmazon-Version 6.2.1](#)
- [EMRAmazon-Version 6.2.0](#)
- [EMRAmazon-Version 6.1.1](#)
- [EMRAmazon-Version 6.1.0](#)
- [EMRAmazon-Version 6.0.1](#)
- [EMRAmazon-Version 6.0.0](#)

Anwendungsversionen in Amazon-EMR-6.x-Versionen

Eine umfassende Tabelle mit den in jeder Amazon-EMR-6.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#) in Ihrem Browser.

EMRAmazon-Version 6.15.0

6.15.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
AWS SDK for 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

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.14.0-amzn-0	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.4.0-amzn-0	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,283	0,281	0,281	0,281
Spark	3.4.1	3.4.1	3.4.1	3.4.0
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

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Tez	0,102	0.10.2	0.10.2	0.10.2
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

6.15.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.15.0. Änderungen beziehen sich auf Version 6.14.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.15.0 Änderungsprotokoll](#).

Neue Features

- Anwendungs-Upgrades – 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.
- [Schnellere Starts für EMR Cluster, die auf einem Cluster laufen EC2](#) — Es ist jetzt bis zu 35% schneller, einen EMR EC2 Amazon-On-Cluster zu starten. Mit dieser Verbesserung können die meisten Kunden ihre Cluster in 5 Minuten oder weniger starten.
- [CodeWhisperer für EMR Studio](#) — Sie können Amazon jetzt CodeWhisperer mit Amazon EMR Studio verwenden, um Empfehlungen in Echtzeit zu erhalten, während Sie Code einschreiben JupyterLab. CodeWhisperer kann Ihre Kommentare vervollständigen, einzelne Codezeilen fertigstellen, line-by-line Empfehlungen aussprechen und vollständig formatierte Funktionen generieren.
- [Schnellere Job-Neustartzeiten mit Flink](#) — Mit Amazon EMR 6.15.0 und höher sind mehrere neue Mechanismen für Apache Flink verfügbar, um die Job-Neustartzeit bei Aufgabenwiederherstellungs- oder Skalierungsvorgängen zu verbessern. Dadurch wird die Geschwindigkeit der Wiederherstellung und des Neustarts von Ausführungsdiagrammen optimiert, um die Auftragsstabilität zu verbessern.
- Detaillierte [Zugriffskontrolle auf Tabellenebene für Open-Table-Formate](#) — Wenn Sie mit Amazon EMR 6.15.0 und höher Spark-Jobs auf Amazon auf EC2 Clustern ausführen, die auf Daten im AWS Glue-Datenkatalog zugreifen, können Sie damit Berechtigungen EMR auf Tabellen-, Zeilen-,

Spalten- und Zellenebene auf Hudi-, Iceberg- oder Delta Lake-basierte Tabellen anwenden. AWS Lake Formation

- Hadoop-Upgrade — Amazon EMR 6.15.0 beinhaltet ein Upgrade von Apache Hadoop auf Version 3.3.6. Hadoop 3.3.6 war die neueste Version zum Zeitpunkt der Bereitstellung von Amazon EMR 6.15, die im Juni 2023 von Apache veröffentlicht wurde. Frühere Versionen von Amazon EMR (6.9.0 bis 6.14.x) verwendeten Hadoop 3.3.3.

Das Upgrade umfasst Hunderte von Verbesserungen und Fehlerbehebungen sowie Funktionen wie rekonfigurierbare Datenknotenparameter, die DFSAdmin Option zum Initiieren von Massenrekonfigurationsvorgängen an allen Live-Datenknoten und eine API Vektorfunktion, die es leseintensiven Lesern ermöglicht, mehrere Lesebereiche anzugeben. Hadoop 3.3.6 bietet außerdem Unterstützung HDFS APIs und Semantik für sein Write-Ahead-Protokoll (`WAL`), sodass es auf anderen Speichersystemimplementierungen ausgeführt werden kann. Weitere Informationen finden Sie in den Änderungsprotokollen für Versionen [3.3.4](#), [3.3.5](#) und [3.3.6](#) in der Apache-Hadoop-Dokumentation.

- Support AWS SDK für Java, Version 2 — Amazon EMR 6.15.0-Anwendungen können AWS SDK für die Java-Versionen [1.12.569](#) oder [2.20.160](#) verwendet werden, wenn die Anwendung v2 unterstützt. Die AWS SDK für Java 2.x ist eine grundlegende Neufassung der Codebasis von Version 1.x. Sie basiert auf Java 8+ und fügt mehrere häufig angeforderte Funktionen hinzu. Dazu gehören die Unterstützung für blockierungsfreie I/O und die Möglichkeit, zur Laufzeit eine andere HTTP Implementierung einzubinden. Weitere Informationen, einschließlich eines Migrationshandbuchs von SDK Java v1 zu v2, finden Sie im Handbuch [AWS SDK für Java, Version 2](#).

Änderungen, Verbesserungen und behobene Probleme

- Um Ihre EMR Hochverfügbarkeitscluster zu verbessern, ermöglicht diese Version Konnektivität zu EMR Amazon-Daemons auf lokalen Hosts, die Endpunkte verwenden. IPv6
- Diese Version ermöglicht TLS 1.2 für die Kommunikation mit ZooKeeper Provisioned auf allen primären Knoten Ihres Hochverfügbarkeitsclusters.
- Diese Version verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und den Clusterbetrieb unterbrechen.
- Diese Version macht die knoteninterne Kommunikation für Hochverfügbarkeitscluster widerstandsfähiger. EMR Durch diese Verbesserung wird die Wahrscheinlichkeit, dass Bootstrap-Aktionen oder Cluster-Startfehler auftreten, verringert.

- Tez in Amazon EMR 6.15.0 führt Konfigurationen ein, die Sie angeben können, um die Eingabe-Splits asynchron in einem Tez-gruppierter Split zu öffnen. Dies führt zu einer schnelleren Leistung von Leseabfragen, wenn ein einzelner gruppierter Tez-Split eine große Anzahl von Eingabe-Splits enthält. Weitere Informationen finden Sie unter [Asynchrone Öffnung von Tez-Splits](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amaon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	13. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

6.15.0 Standard-Java-Versionen

EMRAmazon-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino EMR unterstützt Amazon ab Amazon-Version 6.9.0 standardmäßig Amazon EMR Corretto 17. Amazon unterstützt EMR auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standardeinstellung in JVM Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.15.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
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
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

6.15.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
<code>delta-standalone-connectors</code>	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
<code>emr-ddb</code>	5.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.8.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.12.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.29.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.8.0	EMRS3Select-Anschluss
emr-wal-cli	1.2.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.60.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.1-amzn-1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.1-amzn-1	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.6-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.6-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.6-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.6-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.6-amzn-1	HTTP-Endpoint für Operationen auf HDFS.
hadoop-kms-server	3.3.6-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.

Komponente	Version	Beschreibung
hadoop-mapred	3.3.6-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.6-amzn-1	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.6-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.6-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.4.17-amzn-3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.17-amzn-3	HBase Befehlszeilenclient.
hbase-rest-server	2.4.17-amzn-3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.17-amzn-3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hbase-operator-tools	2.4.17-amzn-3	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-8	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-8	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.3-amzn-8	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-8	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-8	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.14.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.

Komponente	Version	Beschreibung
hudi-presto	0.14.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.14.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.14.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.4.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.283-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.283-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.283-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	426-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.

Komponente	Version	Beschreibung
trino-worker	426-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	426-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.1-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.1-amzn-2	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.4.1-amzn-2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.08.1-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-6	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-6	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTPApache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Client.

6.15.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.15.0 Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-lakeformat ion	Ändern Sie die Werte in der <code>lakeformation.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.15.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.15.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-11-17	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.15.0
17.11.2023-	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.15.0 vollständig in allen unterstützten Regionen bereitgestellt

Datum	Ereignis	Beschreibung
2023-11-13	Erstversion	Amazon EMR 6.15.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 6.14.0

6.14.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
AWS SDK for 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

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
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
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

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
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	0,102	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

6.14.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.14.0. Änderungen beziehen sich auf Version 6.13.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.14.0 Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- [Amazon EMR Managed Scaling](#) ist jetzt in der Region ap-southeast-3 Asien-Pazifik (Jakarta) für Cluster verfügbar, die Sie mit Amazon EMR 6.14.0 und höher erstellen.

Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.14.0 optimiert die Protokollverwaltung, wenn Amazon auf Amazon EMR läuft. EC2 Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Die Version 6.14.0 verbessert den Skalierungs-Workflow, um verschiedene Core-Instances zu berücksichtigen, deren Größe für ihre EBS Amazon-Volumes erheblich variiert. Diese

Verbesserung gilt nur für Core-Knoten; Herunterskalierungs-Operationen für Aufgabenknoten sind davon nicht betroffen.

- Die Version 6.14.0 verbessert die Art und Weise, wie Amazon mit Open-Source-Anwendungen wie EMR interagiert. Apache Hadoop YARN ResourceManager and HDFS NameNode Diese Verbesserung reduziert das Risiko von Betriebsverzögerungen bei der Cluster-Skalierung und verringert Startup-Fehler, die aufgrund von Verbindungsproblemen mit den Open-Source-Anwendungen auftreten.
- Die Version 6.14.0 optimiert die Anwendungsinstallation beim Clusterstart. Dies verbessert die Cluster-Startzeiten für bestimmte Kombinationen von EMR Amazon-Anwendungen.
- Die Version 6.14.0 behebt ein Problem, bei dem Cluster-Scale-Down-Operationen zum Stillstand kommen können, wenn ein Cluster, der in einer VPC mit einer benutzerdefinierten Domain ausgeführt wird, auf einen Core- oder Task-Node-Neustart stößt.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	17. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	11. September 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

6.14.0 Standard-Java-Versionen

EMRAmazon-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino EMR unterstützt Amazon ab Amazon-Version 6.9.0 standardmäßig Amazon EMR Corretto 17. Amazon unterstützt EMR auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standardeinstellung in JVM Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.14.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
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

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

6.14.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.7.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.11.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.28.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.7.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.59.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.1-amzn-0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.1-amzn-0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-6	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-6	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-6	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-6	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-6	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.3.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.3.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-6	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-6	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.17-amzn-2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.17-amzn-2	HBase Befehlszeilenclient.
hbase-rest-server	2.4.17-amzn-2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.17-amzn-2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-2	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-7	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-7	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	3.1.3-amzn-7	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-7	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-7	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.1-amzn-2	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.1-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.1-amzn-2	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.1-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.3.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.281-amzn-2	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.281-amzn-2	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.281-amzn-2	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	422-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	422-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	422-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.4.1-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.1-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.1-amzn-1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.4.1-amzn-1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.06.0-amzn-2	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-5	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-5	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.14.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.14.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der <code>lakeformation.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.14.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.14.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-11-02	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.14.0 vollständig in allen unterstützten Regionen bereitgestellt
2023-10-10	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.14.0

Datum	Ereignis	Beschreibung
2023-10-04	Erstversion	Amazon EMR 6.14.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 6.13.0

6.13.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
AWS SDK for 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

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
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
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

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
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	0,102	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

6.13.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.13.0.

Änderungen beziehen sich auf Version 6.12.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.13.0 Änderungsprotokoll](#).

Neue Features

- 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.

Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.13.0 verbessert den Amazon EMR Log Management Daemon, um sicherzustellen, dass alle Protokolle in regelmäßigen Abständen auf Amazon S3 hochgeladen werden, wenn ein Befehl zur Clusterbeendigung ausgegeben wird. Dies ermöglicht schnellere Clusterbeendigungen.
- Die Version 6.13.0 erweitert die EMR Amazon-Protokollverwaltungsfunktionen, um einen konsistenten und zeitnahen Upload aller Protokolldateien auf Amazon S3 sicherzustellen. Davon profitieren vor allem Cluster mit langer LaufzeitEMR.

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906.0		04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

6.13.0 Standard-Java-Versionen

EMRAmazon-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino EMR unterstützt Amazon ab Amazon-Version 6.9.0 standardmäßig Amazon EMR Corretto 17. Amazon unterstützt EMR auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standardeinstellung in JVM Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.13.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
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

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

6.13.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.6.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.10.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.27.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.6.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.58.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-5	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-5	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-5	HTTP-Endpunkt für Operationen mit HDFS.
hadoop-kms-server	3.3.3-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.3.3-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-5	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-5	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.17-amzn-1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.17-amzn-1	HBase Befehlszeilenclient.
hbase-rest-server	2.4.17-amzn-1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.17-amzn-1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-1	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-6	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-6	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	3.1.3-amzn-6	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-6	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-6	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.1-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.3.0-amzn-1	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.281-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.281-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.281-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	414-amzn-1	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	414-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
trino-client	414-amzn-1	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.4.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.1-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.4.1-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.06.0-amzn-1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-4	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-4	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.13.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.13.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-lakeformat ion	Ändern Sie die Werte in der <code>lakeformation.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.13.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.13.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-09-23	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.13.0 vollständig in allen unterstützten Regionen bereitgestellt
2023-09-12	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.13.0

Datum	Ereignis	Beschreibung
2023-09-01	Erstversion	Amazon EMR 6.13.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 6.12.0

6.12.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
AWS SDK for 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

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
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
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

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
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	0,102	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

6.12.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.12.0.

Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.12.0 Änderungsprotokoll](#).

Neue Features

- 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.
- Amazon EMR veröffentlicht 6.12.0 und höher und unterstützt die LDAP Integration mit Apache Livy, Apache Hive bis HiveServer 2 (HS2), Trino, Presto und Hue. Sie können Apache Spark und Apache Hadoop auch auf einem EMR Cluster installieren, der 6.12.0 oder höher verwendet, und sie für die Verwendung konfigurieren. LDAP Weitere Informationen finden Sie unter [Verwenden von Active Directory oder LDAP Servern für die Authentifizierung bei Amazon EMR](#).

Änderungen, Verbesserungen und behobene Probleme

- EMRAmazon-Versionen 6.12.0 und höher bieten Java 11-Laufzeitunterstützung für Flink. Weitere Informationen finden Sie unter [Flink für die Ausführung mit Java 11 konfigurieren](#).
- Die Version 6.12.0 fügt dem Cluster-Skalierungs-Workflow für EMR Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Die Version 6.12.0 behebt ein Problem, bei dem Cluster-Herunterskalierungs-Operationen zum Stillstand kommen können, wenn ein Core-Knoten, der einer ordnungsgemäßen Außerbetriebnahme unterzogen wird, aus irgendeinem Grund nicht mehr funktionsfähig ist, bevor er vollständig außer Betrieb genommen wird.
- Die Version 6.12.0 verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter den für den Cluster eingestellten Replikationsfaktor herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Die Version 6.12.0 verbessert die Leistung und Effizienz des Health Monitoring Service für Amazon, EMR indem die Geschwindigkeit erhöht wird, mit der Statusänderungen für Instances protokolliert werden. Diese Verbesserung verringert das Risiko einer Leistungseinbuße bei Cluster-Knoten, auf denen mehrere benutzerdefinierte Client-Tools oder Drittanbieteranwendungen ausgeführt werden.
- Die Version 6.12.0 verbessert die Leistung des On-Cluster-Log-Management-Daemons für Amazon. EMR Daher ist die Wahrscheinlichkeit eines Leistungseinbruchs bei EMR Clustern, die Schritte mit hoher Parallelität ausführen, geringer.
- Mit der EMR Amazon-Version 6.12.0 wurde der Log-Management-Daemon aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.
- Die Version 6.12.0 beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.

- Die Version 6.12.0 ermöglicht die Protokollrotation für Timeline Server-Protokolle. YARN Dadurch werden Szenarien mit übermäßiger Festplattenauslastung minimiert, insbesondere bei Clustern mit langer Laufzeit.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amaon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906.0		04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

6.12.0 Standard-Java-Versionen

EMRAmazon-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino EMR unterstützt Amazon ab Amazon-Version 6.9.0 standardmäßig Amazon EMR Corretto 17. Amazon unterstützt EMR auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standardeinstellung in JVM Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon unterstützt EMR nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.12.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
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

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Funke RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

6.12.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.9.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.26.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.5.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.57.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-4	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-4	HTTP-Endpunkt für Operationen mit HDFS.
hadoop-kms-server	3.3.3-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.3.3-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-4	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.17-amzn-0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.17-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.17-amzn-0	HBase Befehlszeilenclient.
hbase-rest-server	2.4.17-amzn-0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.17-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-0	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-5	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-5	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	3.1.3-amzn-5	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-5	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-5	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.3.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.281-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.281-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.281-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	414-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	414-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	414-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.4.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.0-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.4.0-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.06.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-3	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-3	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.12.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.12.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in 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	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Ändern Sie das Java-Home von KMS Hadoop	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-lakeformat ion	Ändern Sie die Werte in der <code>lakeformation.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.12.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.12.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-07-27	Dokumentation der Aktualisierung	Aktualisieren Sie die Java-Optionen für 6.12 und fügen Sie das Oozie-Tutorial zum Update hinzu JVM
2023-07-21	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.12.0 vollständig in allen unterstützten Regionen bereitgestellt

Datum	Ereignis	Beschreibung
2023-07-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.12.0
2023-07-12	Erstversion	Amazon EMR 6.12.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 6.11.1

6.11.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
AWS SDK for Java	1.12.446	1.12,446	1.12,397	1,12,397

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	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
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

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
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	0,279	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	0,102	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

6.11.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.11.1. Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.11.1 Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Aufgrund von Sperrkonflikten kann ein Knoten in einen Deadlock geraten, wenn er gleichzeitig mit dem Versuch, ihn außer Betrieb zu nehmen, hinzugefügt oder entfernt wird. Infolgedessen reagiert der Hadoop Resource Manager (YARN) nicht mehr und wirkt sich auf alle eingehenden und aktuell laufenden Container aus.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version enthält Sicherheitsupdates für Hue und HBase

- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon ausführen, EMR möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith` und `endsWith` erhalten. Like Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn keine benutzerdefinierten Funktionen (UDF) vorhanden sind.
- Diese Version behebt ein Problem, bei dem Containerprotokolle vom Node Log Aggregation Service gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme an S3 weiterleiten kann. YARN
- Diese Version behebt ein Problem mit FairShare Scheduler-Metriken, wenn Node Label für Hadoop aktiviert ist.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version erweitert den Cluster-Skalierungs-Workflow für EMR Cluster, auf denen Presto oder Trino ausgeführt wird, um einen neuen Wiederholungsmechanismus. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter die für den Cluster eingestellte Replikationsfaktor-Einstellung herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 906.0	4,14.322	04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

6.11.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
<code>delta-standalone-connectors</code>	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
<code>emr-ddb</code>	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.8.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.25.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.4.0	EMRS3Select-Anschluss
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.56.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	3.3.3-amzn-3.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-3.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-3.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-3.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-3.1	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	3.3.3-amzn-3.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.3.3-amzn-3.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	3.3.3-amzn-3.1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.15-amzn-1.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.15-amzn-1.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.15-amzn-1.1	HBase-Befehlszeilenclient.
hbase-rest-server	2.4.15-amzn-1.1	Dienst, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.15-amzn-1.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-1.1	Reparaturtool für HBase-Apache-Cluster.
hcatalog-client	3.1.3-amzn-4.1	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.3-amzn-4.1	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-4.1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.3-amzn-4.1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-4.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-4.1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-4.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.2.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.279-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.279-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.279-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	410-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	410-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	410-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.2-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.2-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.2-amzn-0.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.3.2-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.02.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.

Komponente	Version	Beschreibung
tez-on-yarn	0.10.2-amzn-2.1	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-2.1	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTPApache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.11.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.11.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der <code>delta.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.11.1 Änderungsprotokoll

Änderungsprotokoll für Version 6.11.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.11.1
2023-08-16	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.11.1 vollständig in allen unterstützten Regionen bereitgestellt
2023-08-04	Erstversion	Amazon EMR 6.11.1 wurde zuerst in begrenzte

Datum	Ereignis	Beschreibung
		n kommerziellen Regionen bereitgestellt

EMRAmazon-Version 6.11.0

6.11.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
AWS SDK for 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	-	-	-	-

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
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
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

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.10.0
Tez	0,102	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

6.11.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.11.0. Änderungen beziehen sich auf Version 6.10.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.11.0 unterstützt Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg 1.2.0-amzn-0, Trino 410-amzn-0 und PrestoDB 0.279-amzn-0.

Änderungen, Verbesserungen und behobene Probleme

- Mit Amazon EMR 6.11.0 wurde der DynamoDB-Connector auf Version 5.0.0 aktualisiert. Version 5.0.0 verwendet. AWS SDK for Java 2.x In früheren Versionen wurde AWS SDK for Java 1.x verwendet. Aufgrund dieses Upgrades empfehlen wir Ihnen dringend, Ihren Code zu testen, bevor Sie den DynamoDB-Connector mit Amazon EMR 6.11 verwenden.
- Wenn der DynamoDB-Connector für Amazon EMR 6.11.0 den DynamoDB-Service aufruft, verwendet er den Region-Wert, den Sie für die Eigenschaft angeben. `dynamodb.endpoint` Wir empfehlen, dass Sie auch `dynamodb.region` konfigurieren, wenn Sie `dynamodb.endpoint` verwenden, und dass beide Eigenschaften auf dieselbe AWS-Region abzielen. Wenn Sie verwenden `dynamodb.endpoint` und nicht konfigurieren `dynamodb.region`, gibt

der DynamoDB-Connector für Amazon EMR 6.11.0 eine ungültige Region-Ausnahme zurück und versucht, Ihre AWS-Region Informationen aus dem EC2 Amazon-Instance-Metadaten-Service (IMDS) abzugleichen. Wenn der Connector die Region nicht abrufen kann, verwendet er standardmäßig USA Ost (Nord-Virginia) (us-east-1). Der folgende Fehler ist ein Beispiel für die ungültige Region-Ausnahme, die möglicherweise angezeigt wird, wenn Sie die `dynamodb.region` Eigenschaft nicht richtig konfigurieren: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region`. Weitere Informationen zu den Klassen, die vom AWS SDK for Java Upgrade auf 2.x betroffen sind, finden Sie im Commit [Upgrade AWS SDK for Java from 1.x auf 2.x \(#175\)](#) im GitHub Repository für den EMR Amazon-DynamoDB-Connector.

- Diese Version behebt ein Problem, bei dem Spaltendaten zu NULL werden, wenn Sie Delta Lake zum Speichern von Delta-Tabellendaten in Amazon S3 nach dem Spaltenumbenennungsvorgang verwenden. Weitere Informationen zu diesem experimentellen Feature in Delta Lake finden Sie unter [Vorgang zum Umbenennen von Spalten](#) im Delta-Lake-Benutzerhandbuch.
- Die Version 6.11.0 behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen eines Edge-Knotens für die Kommunikation mit Ihrem EMR Cluster finden Sie im Repo unter [Edge Node](#) Creator. `aws-samples` GitHub
- Die Version 6.11.0 verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um EBS Amazon-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Die Version 6.11.0 behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon auf Amazon EMR veröffentlicht. `CloudWatch`
- Die Version 6.11.0 behebt ein Problem mit EMR Clustern, bei dem eine Aktualisierung der YARN Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer übermäßigen Festplattenauslastung unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container ausgeführt wurden, so lange außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie

Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` ist der Wert von `true` auf `false` in `yarn-site.xml` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN -9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Jobs aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit EMR Amazon-Versionen 6.11.0 und höher verfügbar.

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Architektur. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI-Aktion [ListReleaseLabels](#) API-Operation. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Arta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504,1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

6.11.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
<code>delta-standalone-connectors</code>	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
<code>emr-ddb</code>	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.8.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.25.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.4.0	EMRS3Select-Anschluss
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.56.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-3	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-3	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-3	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.3.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	3.3.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-3	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-3	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.4.15-amzn-1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.15-amzn-1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.15-amzn-1	HBaseBefehlszeilenclient.
hbase-rest-server	2.4.15-amzn-1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.15-amzn-1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hbase-operator-tools	2.4.15-amzn-1	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-4	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-4	HTTP-Endpoint, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.3-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.

Komponente	Version	Beschreibung
hudi-presto	0.13.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.2.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.279-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.279-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.279-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	410-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.

Komponente	Version	Beschreibung
trino-worker	410-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	410-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.2-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.2-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.2-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.3.2-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	23.02.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mitGPUs.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-2	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-2	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTPApache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Client.

6.11.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.11.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der <code>delta.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.11.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.11.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-21	Aktualisierung	Das mit Hadoop 3.3.3 eingeführte Problem wurde behoben.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
2023-06-09	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.11.0 vollständig in allen unterstützten Regionen bereitgestellt
2023-06-09	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.11.0

Datum	Ereignis	Beschreibung
2023-06-08	Erstversion	Amazon EMR 6.11.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

EMRAmazon-Version 6.10.1

6.10.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
AWS SDK for 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

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.1.0	2.1.0
Flink	1.16.0	1.16.0	1.15.2	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.13	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
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

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
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	0,102	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

6.10.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.10.1.

Änderungen beziehen sich auf Version 6.10.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.10.1 Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Aufgrund von Sperrkonflikten kann ein Knoten in einen Deadlock geraten, wenn er gleichzeitig mit dem Versuch, ihn außer Betrieb zu nehmen, hinzugefügt oder entfernt wird. Infolgedessen reagiert der Hadoop Resource Manager (YARN) nicht mehr und wirkt sich auf alle eingehenden und aktuell laufenden Container aus.
- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container ausgeführt wurden, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` ist der Wert von auf `false` in `yarn-site.xml` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN -9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Jobs aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit EMR Amazon-Versionen 6.11.0 und höher verfügbar.

- Der Metrics Collector sendet nach einem Failover des Primärknotens in Clustern mit der Instance-Gruppenkonfiguration keine Metriken an die Steuerebene.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version enthält Sicherheitskorrekturen für Hue und HBase
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon ausführen, EMR möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith` und `endsWith` erhalten. Like Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server () haben. HMS
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn keine benutzerdefinierten Funktionen () vorhanden sind. UDF
- Diese Version behebt ein Problem, bei dem Containerprotokolle vom Node Log Aggregation Service gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme an S3 weiterleiten kann. YARN
- Diese Version behebt ein Problem mit FairShare Scheduler-Metriken, wenn Node Label für Hadoop aktiviert ist.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version erweitert den Cluster-Skalierungs-Workflow für EMR Cluster, auf denen Presto oder Trino ausgeführt wird, um einen neuen Wiederholungsmechanismus. Diese Verbesserung

verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.

- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter die für den Cluster eingestellte Replikationsfaktor-Einstellung herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Diese Version behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen eines Edge-Knotens für die Kommunikation mit Ihrem EMR Cluster finden Sie unter [Edge Node Creator](#) im Repo unter `aws-samples` GitHub
- Diese Version verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um EBS Amazon-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Diese Version behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon auf Amazon EMR veröffentlicht. CloudWatch
- Diese Version behebt ein Problem mit EMR Clustern, bei dem eine Aktualisierung der YARN Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer übermäßigen Festplattenauslastung unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906.0		04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

6.10.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.7.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.24.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.3.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emr-wal-cli	1.0.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.55.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-2.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-2.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-2.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-2.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-https-server	3.3.3-amzn-2.1	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	3.3.3-amzn-2.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.3.3-amzn-2.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-2.1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.15-amzn-0.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.15-amzn-0.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.15-amzn-0.1	HBase Befehlszeilenclient.
hbase-rest-server	2.4.15-amzn-0.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.15-amzn-0.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-0.1	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-3.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-3.1	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-3.1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	3.1.3-amzn-3.1	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-3.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-3.1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-3.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.2-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.2-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.1.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0- SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	6.0.0	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.278.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.278.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.278.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	403-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	403-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	403-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.3.1-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.1-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.1-amzn-0.1	In-Memory-Ausführung engine für YARN
spark-yarn-slave	3.3.1-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.12.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-1.1	Die YARN TEZ-Anwendung und die Bibliotheken.
tez-on-worker	0.10.2-amzn-1.1	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.10.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.10.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die <code>flink-conf.yaml</code> -Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink <code>log4j-session.properties</code> für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der <code>delta.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.10.1 Änderungsprotokoll

Änderungsprotokoll für Version 6.10.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.10.1
2023-08-16	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.10.1 vollständig in allen unterstützten Regionen bereitgestellt
2023-08-04	Erstversion	Amazon EMR 6.10.1 wurde zuerst in begrenzte

Datum	Ereignis	Beschreibung
		n kommerziellen Regionen bereitgestellt

EMRAmazon-Version 6.10.0

6.10.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
AWS SDK for 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	-	-	-	-

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
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
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

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
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

6.10.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.10.0.

Änderungen beziehen sich auf Version 6.9.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.10.0 unterstützt 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 und PrestoDB 0.278.1.
- Amazon EMR 6.10.0 enthält einen nativen Trino-Hudi-Konnektor, der Lesezugriff auf Daten in Hudi-Tabellen ermöglicht. Sie können den Konnektor mit `trino-cli --catalog hudi` aktivieren und den Konnektor für Ihre Anforderungen mit `trino-connector-hudi` konfigurieren. EMR Durch die native Integration mit Amazon müssen Sie Hudi-Tabellen nicht mehr abfragen. `trino-connector-hive` Eine Liste der unterstützten Konfigurationen mit dem neuen Konnektor finden Sie auf der [Hudi-Konnektor-Seite](#) der Trino-Dokumentation.
- EMR Amazon-Versionen 6.10.0 und höher unterstützen die Apache Zeppelin-Integration mit Apache Flink. Weitere Informationen finden Sie unter [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#).

Bekannte Probleme

- Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Um dieses Problem in Amazon EMR 6.10.0 zu umgehen, können Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` setzen. In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration `false` standardmäßig auf eingestellt, um dieses Problem zu beheben.

Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR 6.10.0 entfernt die Abhängigkeit von der [Amazon Redshift Redshift-Integration minimal-json.jar für Apache Spark](#) und fügt automatisch die erforderlichen Spark-Redshift-bezogenen JAR-Dateien zum Executor-Klassenpfad für Spark hinzu: `und. spark-redshift.jar` `spark-avro.jar` `RedshiftJDBC.jar`
- Die Version 6.10.0 verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem Cluster. EMR Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Die Version 6.10.0 startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Amazon EMR 6.10.0 unterstützt regionale Endpunkte für die EMRFS Benutzerzuweisung.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Die Version 6.10.0 behebt ein Problem, das dazu führte, dass Spark-Jobs zum Stillstand kamen, wenn sich alle verbleibenden Spark-Executoren auf einem außer Betrieb genommenen Host mit dem Resource Manager befanden. YARN
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload

keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf `-1` zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (`6.8.1`) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMR Amazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API oder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			<p>Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)</p>

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Arta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504,1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Arta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Arta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_404,1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)
2.0.2023_404,0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

6.10.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich

EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.7.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
<code>emr-s3-dist-cp</code>	2.24.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	2.3.0	EMRS3Select-Anschluss
emr-wal-cli	1.0.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.55.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	3.3.3-amzn-2	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.3.3-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-2	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-2	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.3.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.3.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-2	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-2	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.4.15-amzn-0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.15-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.15-amzn-0	HBase Befehlszeilenclient.
hbase-rest-server	2.4.15-amzn-0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.15-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-0	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-3	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.3-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-3	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.2-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.2-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.1.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0- SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	6.0.0	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.278.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.278.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.278.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	403-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	403-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	403-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.1-amzn-0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	3.3.1-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.12.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-1	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-1	Die YARN TEZ-Anwendung und die Bibliotheken für Worker-Knoten.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.10.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.10.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der <code>delta.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der Datei kms-logback.xml von KMS Ranger.	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine SQL SSL Verbindung mit KMS Ranger.	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.10.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.10.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-21	Aktualisierung	Es wurde ein bekanntes Problem hinzugefügt, das mit Hadoop 3.3.3 eingefügt wurde.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
2023-03-02	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.10 vollständig in allen unterstützten Regionen bereitgestellt
2023-03-02	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.10

Datum	Ereignis	Beschreibung
2023-02-27	Erstversion	Amazon EMR 6.10 wird in begrenzten kommerziellen Regionen bereitgestellt

EMRAmazon-Version 6.9.1

6.9.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
AWS SDK for 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

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
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
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

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
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

6.9.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.9.1. Änderungen beziehen sich auf Version 6.9.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.9.1 Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` ist der Wert von `true` auf `false` in `yarn-site.xml` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN -9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Jobs aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem

wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit EMR Amazon-Versionen 6.11.0 und höher verfügbar.

- Der Metrics Collector sendet nach einem Failover des Primärknotens in Clustern mit der Instance-Gruppenkonfiguration keine Metriken an die Steuerebene.
- In dieser Version werden Wiederholungsversuche bei fehlgeschlagenen HTTP Anfragen an Endpunkte zur Erfassung von Metriken vermieden.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version behebt ein Problem, bei dem große, von Benutzern erstellte Ausnahmen zu Überläufen UIDs führten.
- Diese Version behebt Timeout-Probleme beim EMR Amazon-Rekonfigurationsprozess.
- Diese Version enthält Sicherheitskorrekturen.
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon ausführen, EMR möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith` und `endsWith` erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR
- Bei Hive kann es zu Datenverlusten kommen, wenn Sie es HDFS als Scratch-Verzeichnis verwenden und die Option „Kleine Dateien zusammenführen“ aktiviert haben und die Tabelle statische Partitionspfade enthält.
- Diese Version behebt ein Leistungsproblem mit Hive, wenn das Zusammenführen kleiner Dateien (standardmäßig deaktiviert) am Ende des ETL Jobs aktiviert ist.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn keine benutzerdefinierten Funktionen (UDF) vorhanden sind.

- Diese Version behebt ein Problem, bei dem Containerprotokolle vom Node Log Aggregation Service gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme an S3 weiterleiten kann. YARN
- Diese Version behebt den Umgang mit komprimierten/archivierten Dateien mit persistenter Speicherdateiverfolgung für. HBase
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version behebt ein Problem, das dazu führte, dass der Node Provisioner ausfiel, wenn der Dienst HDFS NameNode (NN) während des Knotenaustauschs im Safemode hängen blieb.
- Diese Version erweitert den Cluster-Skalierungsablauf für EMR Cluster, auf denen Presto oder Trino ausgeführt wird, um einen neuen Wiederholungsmechanismus. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter die für den Cluster eingestellte Replikationsfaktor-Einstellung herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Diese Version behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen

eines Edge-Knotens für die Kommunikation mit Ihrem EMR Cluster finden Sie unter [Edge Node Creator](#) im Repo unter `aws-samples` GitHub

- Diese Version verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um EBS Amazon-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Diese Version behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon auf Amazon EMR veröffentlicht. CloudWatch
- Diese Version behebt ein Problem mit EMR Clustern, bei dem eine Aktualisierung der YARN Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer übermäßigen Festplattenauslastung unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Diese Version verbessert den Log-Management-Daemon auf dem Cluster, sodass zusätzliche Protokollordner in Ihrem Cluster überwacht werden können. EMR Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Diese Version startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906,0		04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

6.9.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.6.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.23.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.2.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.54.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.15.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.2	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-1.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-1.1	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-1.1	HDFS Befehlszeilenclient und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.3.3-amzn-1.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-1.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-1.1	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.3.3-amzn-1.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.3.3-amzn-1.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-1.1	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-1.1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-1.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.13-amzn-0.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.13-amzn-0.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.13-amzn-0.1	HBase Befehlszeilenclient.
hbase-rest-server	2.4.13-amzn-0.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.13-amzn-0.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.13-amzn-0.1	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-2.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-2.1	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-2.1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	3.1.3-amzn-2.1	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-2.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-2.1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-2.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.14.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0- SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	6.0.0	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.276-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.276-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.276-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	398-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	398-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	398-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.3.0-amzn-1.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-1.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	3.3.0-amzn-1.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.08.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mitGPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.10.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-0.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTPApache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.9.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.9.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceMan-anager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der <code>delta.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.9.1 Änderungsprotokoll

Änderungsprotokoll für Version 6.9.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.9.1
2023-08-16	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.9.1 vollständig in allen unterstützten Regionen bereitgestellt
2023-08-04	Erstversion	Amazon EMR 6.9.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

EMRAmazon-Version 6.9.0

6.9.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
AWS SDK for 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

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
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
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

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
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

6.9.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.9.0. Die Änderungen beziehen sich auf EMR Amazon-Version 6.8.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Neue Features

- EMRAmazon-Version 6.9.0 unterstützt Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 und Tez 0.10.2.
- EMRAmazon-Version 6.9.0 enthält eine neue Open-Source-Anwendung, [Delta Lake](#) 2.1.0.
- Die Amazon Redshift Redshift-Integration für Apache Spark ist in EMR Amazon-Versionen 6.9.0 und höher enthalten. Die native Integration war bisher ein Open-Source-Tool und ist ein Spark-Konnektor, mit dem Sie Apache-Spark-Anwendungen erstellen können, die Daten in Amazon Redshift und Amazon Redshift Serverless lesen und in diese schreiben. Weitere Informationen finden Sie unter [Verwenden der Amazon Redshift Redshift-Integration für Apache Spark mit Amazon EMR](#).
- Die EMR Amazon-Version 6.9.0 bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Scale-Down. Bisher konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die neue Funktion stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).
- Um Abfragen mit langer Laufzeit zu unterstützen, verfügt Trino jetzt über einen fehlertoleranten Ausführungsmechanismus. Die fehlertolerante Ausführung minimiert Abfragefehler, indem fehlgeschlagene Abfragen oder deren Komponentenaufgaben wiederholt werden. Weitere Informationen finden Sie unter [Fehlertolerante Ausführung in Trino](#).

- Sie können Apache Flink on Amazon EMR für die Vereinheitlichung BATCH und STREAM Verarbeitung von Apache Hive-Tabellen oder Metadaten beliebiger Flink-Tabellenquellen wie Iceberg, Kinesis oder Kafka verwenden. Sie können den AWS Glue-Datenkatalog als Metastore für Flink angeben, indem Sie Amazon AWS Management Console, AWS CLI, oder verwenden. EMR API Weitere Informationen finden Sie unter [Konfiguration von Flink in Amazon EMR](#).
- Sie können jetzt AWS Identity and Access Management (IAM) Runtime-Rollen und AWS Lake Formation basierte Zugriffskontrolle für Apache Spark-, Apache Hive- und Presto-Abfragen auf Amazon EMR auf EC2 Clustern mit Amazon SageMaker Studio angeben. Weitere Informationen finden [Sie unter EMR Schritte zur Konfiguration von Runtime-Rollen für Amazon](#).

Bekannte Probleme

- Für die EMR Amazon-Version 6.9.0 funktioniert Trino nicht auf Clustern, die für Apache Ranger aktiviert sind. Wenn Sie Trino mit Ranger verwenden müssen, wenden Sie sich an [AWS Support](#).
- Wenn Sie die Amazon-Redshift-Integration für Apache Spark verwenden und eine Zeit, `timetz`, `timestamp` oder `timestampz` mit Mikrosekundengenauigkeit im Parquet-Format haben, rundet der Konnektor die Zeitwerte auf den nächstliegenden Millisekundenwert. Um das Problem zu umgehen, verwenden Sie den `unload_s3_format`-Formatparameter-Text-Unload.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Verbindungen zu EMR Amazon-Clustern von Amazon SageMaker Studio aus können zeitweise fehlschlagen und der Antwortcode 403 Forbidden wird angezeigt. Dieser Fehler tritt auf, wenn die Einrichtung der IAM Rolle im Cluster länger als 60 Sekunden dauert. Um das Problem zu umgehen, können Sie einen EMR Amazon-Patch installieren, um Wiederholungsversuche zu ermöglichen und das Timeout auf mindestens 300 Sekunden zu erhöhen. Verwenden Sie die folgenden Schritte, um die Bootstrap-Aktion anzuwenden, wenn Sie Ihren Cluster starten.

1. Laden Sie das Bootstrap-Skript und die RPM Dateien vom folgenden Amazon S3 URIs herunter.

```
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. Laden Sie die Dateien aus dem vorherigen Schritt in einen Amazon-S3-Bucket hoch, dessen Eigentümer Sie sind. Der Bucket muss sich in dem Bereich befinden AWS-Region , in dem Sie den Cluster starten möchten.
3. Schließen Sie beim Starten Ihres EMR Clusters die folgende Bootstrap-Aktion ein. Ersetzen *bootstrap_URI* and *RPM_URI* mit dem entsprechenden URIs von Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei RecordServer Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. SecretAgent Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink bietet native S3 FileSystem - und FileSystem Hadoop-Konnektoren, mit denen Anwendungen Daten erstellen FileSink und in Amazon S3 schreiben können. Dies FileSink schlägt mit einer der folgenden zwei Ausnahmen fehl.

```
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:?]
```

Als Workaround können Sie einen EMR Amazon-Patch installieren, der das oben genannte Problem in Flink behebt. Führen Sie die folgenden Schritte aus, um die Bootstrap-Aktion beim Start Ihres Clusters anzuwenden.

1. Laden Sie das flink-rpm in Ihren Amazon-S3-Bucket herunter. Ihr RPM Weg ist. `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`
2. Laden Sie das Bootstrap-Skript und die RPM Dateien wie folgt URI von Amazon S3 herunter. `regionName` Ersetzen Sie es durch den AWS-Region Ort, an dem Sie den Cluster starten möchten.

```
s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-rpms.sh
```

3. Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser


Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierte verwaltete Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es eine Lösung für dieses Problem, indem Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` in den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration `false` standardmäßig auf eingestellt, um dieses Problem zu beheben.

Änderungen, Verbesserungen und behobene Probleme

- Für EMR Amazon-Version 6.9.0 und höher verwenden alle von Amazon installierten Komponenten, EMR die Log4j-Bibliotheken verwenden, Log4j Version 2.17.1 oder höher.
- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. EMRAmazon-Version 6.9.0 behebt dieses Problem.
- Amazon EMR 6.9.0 bietet eingeschränkte Unterstützung für die auf Lake Formation basierende Zugriffskontrolle mit Apache Hudi beim Lesen von Daten mit Spark. SQL Die Unterstützung bezieht sich auf SELECT Abfragen, die Spark verwenden, SQL und ist auf die Zugriffskontrolle auf Spaltenebene beschränkt. Weitere Informationen finden Sie unter [Hudi und Lake Formation](#).
- Wenn Sie Amazon EMR 6.9.0 verwenden, um einen Hadoop-Cluster mit aktivierten [Node Labels](#) zu erstellen, geben die [YARNMetriken](#) aggregierte Informationen über alle Partitionen statt der Standardpartition API zurück. [Weitere Informationen finden Sie unter -11414. YARN](#)
- Mit der EMR Amazon-Version 6.9.0 haben wir Trino auf Version 398 aktualisiert, die Java 17 verwendet. Die vorherige unterstützte Version von Trino für Amazon EMR 6.8.0 war Trino 388, das auf Java 11 lief. Weitere Informationen zu dieser Änderung finden Sie unter [Trino-Updates für Java 17 im Trino-Blog](#).
- Diese Version behebt ein Problem mit der Nichtübereinstimmung der Timing-Sequenz zwischen Apache BigTop und EMR Amazon bei der EC2 Cluster-Startsequenz. Diese Nichtübereinstimmung der Timing-Sequenz tritt auf, wenn ein System versucht, zwei oder mehr Operationen gleichzeitig auszuführen, anstatt sie in der richtigen Reihenfolge auszuführen. Infolgedessen kam es bei bestimmten Clusterkonfigurationen zu Timeouts beim Instance-Start und zu langsameren Cluster-Startup-Zeiten.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon

Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

 Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6 . 8 . **1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API Loder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_612.0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	12. Januar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

6.9.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.6.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.23.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.2.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.54.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.15.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.2	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-1	HDFSBefehlszeilenclient und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.3.3-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-1	HTTP-Endpunkt für Operationen mit HDFS.
hadoop-kms-server	3.3.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.3.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-1	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.13-amzn-0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.13-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.13-amzn-0	HBase Befehlszeilenclient.
hbase-rest-server	2.4.13-amzn-0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.13-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.13-amzn-0	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-2	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	3.1.3-amzn-2	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-2	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.14.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0- SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	6.0.0	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.276-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.276-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.276-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	398-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	398-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	398-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.3.0-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.3.0-amzn-1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.08.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.10.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-0	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.9.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.9.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in 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	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die <code>log4j2.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der <code>delta.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.9.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.9.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Korrektur für das Problem mit der Nichtübereinstimmung der Timing-Sequenz hinzugefügt
2023-08-21	Versionshinweise der Aktualisierung	Es wurde ein bekanntes Problem mit Hadoop 3.3.3 hinzugefügt.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
13.12.2022	Versionshinweise	Funktion hinzugefügt und bekanntes Problem für Runtime mit SageMaker

Datum	Ereignis	Beschreibung
29.11.2022	Versionshinweise und Dokumentation wurden aktualisiert	Feature für die Amazon-Redshift-Integration für Apache Spark hinzugefügt
23.11.2022	Versionshinweise	Der Log4j-Eintrag wurde entfernt
18.11.2022	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.9 vollständig in allen unterstützten Regionen bereitgestellt
18.11.2022	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.9
14.11.2022	Erstversion	Amazon EMR 6.9 wird in begrenzten kommerziellen Regionen bereitgestellt

EMRAmazon-Version 6.8.1

6.8.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
AWS SDK for 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	-	-	-	-
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

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
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
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

6.8.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.8.1.

Änderungen beziehen sich auf Version 6.8.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.8.1 Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist.

Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` ist der Wert von `true` auf `false` in `yarn-site.xml` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN -9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Jobs aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit EMR Amazon-Versionen 6.11.0 und höher verfügbar.

- Der Metrics Collector sendet nach einem Failover des Primärknotens in Clustern mit der Instance-Gruppenkonfiguration keine Metriken an die Steuerebene.
- In dieser Version werden Wiederholungsversuche bei fehlgeschlagenen HTTP Anfragen an Endpunkte zur Erfassung von Metriken vermieden.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version behebt ein Problem, bei dem große, von Benutzern erstellte Ausnahmen zu Überläufen UIDs führten.
- Diese Version behebt Timeout-Probleme beim EMR Amazon-Rekonfigurationsprozess.
- Diese Version verhindert ein Problem, bei dem eine fehlgeschlagene Neukonfiguration andere, nicht miteinander verbundene Prozesse unterbrechen könnte.
- Diese Version enthält Sicherheitskorrekturen.
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon ausführen, EMR möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith` und `endsWith` erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload

keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf `-1` zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR

- Bei Hive kann es zu Datenverlusten kommen, wenn Sie es HDFS als Scratch-Verzeichnis verwenden und die Option „Kleine Dateien zusammenführen“ aktiviert haben und die Tabelle statische Partitionspfade enthält.
- Diese Version behebt ein Leistungsproblem mit Hive, wenn das Zusammenführen kleiner Dateien (standardmäßig deaktiviert) am Ende des ETL Jobs aktiviert ist.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn keine benutzerdefinierten Funktionen () vorhanden sind. UDF
- Diese Version behebt ein Problem, bei dem Containerprotokolle vom Node Log Aggregation Service gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme an S3 weiterleiten kann. YARN
- Diese Version behebt den Umgang mit komprimierten/archivierten Dateien mit persistenter Speicherdateiverfolgung für. HBase
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version behebt ein Problem, das dazu führte, dass der Node Provisioner ausfiel, wenn der Dienst HDFS NameNode (NN) während des Knotenaustauschs im Safemode hängen blieb.
- Diese Version erweitert den Cluster-Skalierungsablauf für EMR Cluster, auf denen Presto oder Trino ausgeführt wird, um einen neuen Wiederholungsmechanismus. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter die für den Cluster eingestellte Replikationsfaktor-Einstellung herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen

Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.

- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Diese Version behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen eines Edge-Knotens für die Kommunikation mit Ihrem EMR Cluster finden Sie unter [Edge Node Creator](#) im Repo unter `aws-samples` GitHub
- Diese Version verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um EBS Amazon-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Diese Version behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon auf Amazon EMR veröffentlicht. CloudWatch
- Diese Version behebt ein Problem mit EMR Clustern, bei dem eine Aktualisierung der YARN Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer übermäßigen Festplattenauslastung unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Diese Version verbessert den Log-Management-Daemon auf dem Cluster, sodass zusätzliche Protokollordner in Ihrem Cluster überwacht werden können. EMR Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Diese Version startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Diese Version bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Scale-Down. Bisher konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die neue Funktion stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).

- Diese Version behebt ein Problem, das auftrat, als die Amazon S3 S3-Aktion URI für eine Bootstrap-Aktion mit einer Portnummer endete, zum Beispiel: a . b . c . d : 4345. Amazon EMR hat diese falsch analysiert URIs, sodass alle zugehörigen Bootstrap-Aktionen fehlschlagen würden.
- Diese Version behebt ein Problem mit der Nichtübereinstimmung der Timing-Sequenz zwischen Apache BigTop und EMR Amazon bei der EC2 Cluster-Startsequenz. Diese Nichtübereinstimmung der Timing-Sequenz tritt auf, wenn ein System versucht, zwei oder mehr Operationen gleichzeitig auszuführen, anstatt sie in der richtigen Reihenfolge auszuführen. Infolgedessen kam es bei bestimmten Clusterkonfigurationen zu Timeouts beim Instance-Start und zu langsameren Cluster-Startup-Zeiten.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906,0		04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

6.8.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.22.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.53.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.15.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-8.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-8.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-8.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-8.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-8.1	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.2.1-amzn-8.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-8.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-8.1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.4.12-amzn-0.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	2.4.12-amzn-0.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.12-amzn-0.1	HBase Befehlszeilenclient.
hbase-rest-server	2.4.12-amzn-0.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.12-amzn-0.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.12-amzn-0.1	Reparaturtool für HBase Apache-Cluster.
hcatalog-client	3.1.3-amzn-1.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-1.1	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-1.1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	3.1.3-amzn-1.1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-1.1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	3.1.3-amzn-1.1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-1.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.11.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.11.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.14.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.273.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.273.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.273.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	388-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	388-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	388-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System

Komponente	Version	Beschreibung
spark-client	3.3.0-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-0.1	In-Memory-Ausführung engine für YARN
spark-yarn-slave	3.3.0-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.06.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.9.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.8.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.8.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceMananager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die <code>flink-conf.yaml</code> -Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink <code>log4j-session.properties</code> für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der <code>raptor.properties</code> -Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der <code>redis.properties</code> -Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der <code>redshift.properties</code> -Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der <code>tpch.properties</code> -Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der <code>tpcds.properties</code> -Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei <code>dbks-site.xml</code> von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der <code>ranger-kms-site</code> XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.8.1 Änderungsprotokoll

Änderungsprotokoll für Version 6.8.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.8.1
2023-08-16	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.8.1 vollständig in allen unterstützten Regionen bereitgestellt
2023-08-04	Erstversion	Amazon EMR 6.8.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

EMRAmazon-Version 6.8.0

6.8.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
AWS SDK for Java	1.12.170	1.12.170	1.12.170	1.12,31
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	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

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
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
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

6.8.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.8.0. Änderungen beziehen sich auf Version 6.7.0.

Neue Features

- Die Amazon EMR Steps-Funktion unterstützt jetzt Apache Livy-Endpunkte und JDBC ODBC /- Clients. Weitere Informationen finden [Sie unter EMR Schritte zur Konfiguration von Runtime-Rollen für Amazon](#).
- EMRAmazon-Version 6.8.0 wird mit Apache HBase Version 2.4.12 geliefert. Mit dieser HBase Version können Sie Ihre Tabellen sowohl archivieren als auch löschen. HBase Der Amazon-S3-Archivierungsprozess benennt alle Tabellendateien in das Archivverzeichnis um. Dies kann ein kostspieliger und langwieriger Prozess sein. Jetzt können Sie den Archivierungsvorgang überspringen und große Tabellen schnell löschen und löschen. Weitere Informationen finden Sie unter [Verwenden der HBase Shell](#).

Bekannte Probleme


- Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierte verwaltete Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es eine Lösung für dieses Problem, indem Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration `false` standardmäßig auf eingestellt, um dieses Problem zu beheben.

Änderungen, Verbesserungen und behobene Probleme

- Als Amazon EMR Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell las, produzierte Amazon EMR eine. `NoSuchMethodError` EMRAmazon-Version 6.8.0 behebt dieses Problem.

- Die EMR Amazon-Version 6.8.0 enthält [Apache Hudi](#) 0.11.1. Amazon EMR 6.8.0-Cluster sind jedoch auch mit der Open-Source-Version von Hudi 0.12.0 kompatibel. `hudi-spark3.3-bundle_2.12`
- EMRAmazon-Version 6.8.0 wird mit Apache Spark 3.3.0 geliefert. Diese Spark-Version verwendet Apache Log4j 2 und die `log4j2.properties`-Datei zur Konfiguration von Log4j in Spark-Prozessen. Wenn Sie Spark im Cluster verwenden oder EMR Cluster mit benutzerdefinierten Konfigurationsparametern erstellen und ein Upgrade auf EMR Amazon-Version 6.8.0 durchführen möchten, müssen Sie auf die neue `spark-log4j2` Konfigurationsklassifizierung und das neue Schlüsselformat für Apache Log4j 2 migrieren. Weitere Informationen finden Sie unter [Migration von Apache Log4j 1.x zu Log4j 2.x](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Azon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

 Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (`6.8.1`) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API Loder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv)
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4.14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	6. September 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

Bekannte Probleme

- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem

zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`.
EMRAmazon-Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei RecordServer Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. SecretAgent Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

6.8.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.22.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.53.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.15.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	3.2.1-amzn-8	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-8	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-8	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-8	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-8	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-8	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	3.2.1-amzn-8	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.2.1-amzn-8	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-8	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	3.2.1-amzn-8	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-8	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.12-amzn-0	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.12-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.12-amzn-0	HBase-Befehlszeilenclient.
hbase-rest-server	2.4.12-amzn-0	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.12-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.12-amzn-0	Reparaturtool für HBase-Apache-Cluster.
hcatalog-client	3.1.3-amzn-1	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.3-amzn-1	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.11.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.11.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.14.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.273.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.273.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.273.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	388-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	388-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	388-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.3.0-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.06.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.9.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.

Komponente	Version	Beschreibung
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

6.8.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.8.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der <code>raptor.properties</code> -Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der <code>redis.properties</code> -Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der <code>redshift.properties</code> -Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der <code>tpch.properties</code> -Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der <code>tpcds.properties</code> -Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei <code>dbks-site.xml</code> von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der <code>ranger-kms-site</code> XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

6.8.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.8.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-21	Aktualisierung	Es wurde ein bekanntes Problem mit Hadoop 3.3.3 hinzugefügt.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
06.09.2022	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.8 wurde vollständig in allen unterstützten Regionen bereitgestellt
06.09.2022	Erste Veröffentlichung	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.8
31.08.2022	Erstversion	Amazon EMR 6.8 wurde für begrenzte kommerzielle Regionen veröffentlicht

EMRAmazon-Version 6.7.0

6.7.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
AWS SDK for 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

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
HBase	2.4.4	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.2	3.1.2	3.1.2
Hudi	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	0.13.1-amzn-0	0.13.1	0.12.0	-
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

6.7.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.7.0. Änderungen beziehen sich auf Version 6.6.0.

Erste Version: 15. Juli 2022

Neue Features

- Amazon unterstützt EMR jetzt Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 und Trino 0.378.
- Unterstützt IAM Role- und Lake Formation Formation-basierte Zugriffskontrollen mit EMR Schritten (Spark, Hive) für Amazon EMR auf Clustern. EC2
- Unterstützt Apache-Spark-Datendefinitionsanweisungen auf Apache-Ranger-fähigen Clustern. Dies beinhaltet jetzt die Unterstützung für Trino-Anwendungen, die Apache-Hive-Metadaten auf Apache-Ranger-fähigen Clustern lesen und schreiben. Weitere Informationen finden Sie unter [Aktivieren von Federated Governance mit Trino und Apache Ranger](#) auf Amazon. EMR
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Arbeitsschritt-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 906,0	4,14.322	04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230414.1	4.14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 719.0	4,14.287	10. August 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 606.1	4,14.281	15. Juli 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

Bekannte Probleme

- Wenn EMR Amazon-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, `NoSuchMethodError` tritt ein, weil Amazon ein falsches EMR verwendet. `Hbase.compat.version` EMRAmazon-Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück,

obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. EMRAmazon-Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei RecordServer Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. SecretAgent Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Das wird bei Clustern, die auf Amazon EMR 6.7 oder niedriger ausgeführt werden, GetClusterSessionCredentials API nicht unterstützt.

6.7.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.6.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.22.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.52.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-7	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-7	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-7	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-7	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-7	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-7	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.2.1-amzn-7	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-7	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.2.1-amzn-7	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-7	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-7	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.4-amzn-3	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.4-amzn-3	HBase-Befehlszeilenclient.
hbase-rest-server	2.4.4-amzn-3	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.4-amzn-3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.4-amzn-3	Reparaturtool für HBase Apache-Cluster.

Komponente	Version	Beschreibung
hcatalog-client	3.1.3-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-0	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.3-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltH Catalog.
hive-client	3.1.3-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-0	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.3-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.11.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.11.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hudi-trino	0.11.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.11.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.13.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.272-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.272-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.272-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	378-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	378-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
trino-client	378-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.2.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.2.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.2.1-amzn-0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	3.2.1-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.02.0-amzn-1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Cli-ent.

6.7.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.7.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der <code>raptor.properties</code> -Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der <code>redis.properties</code> -Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der <code>redshift.properties</code> -Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der <code>tpch.properties</code> -Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der <code>tpcds.properties</code> -Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei <code>dbks-site.xml</code> von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der <code>ranger-kms-site</code> XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.6.0

6.6.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
AWS SDK for 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

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.0	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0
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

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Presto	0,267	0,261	0,254,1	0,245,1
Spark	3.2.0	3.1.2	3.1.2	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	367	360	359	350
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3,5.7	3.5.7	3.5.7	3.4.14

6.6.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.6.0. Änderungen beziehen sich auf Version 6.5.0.

Erste Version: 09. Mai 2022

Letzte Aktualisierung der Dokumentation: 15. Juni 2022

Neue Features

- Amazon EMR 6.6 unterstützt jetzt Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 und PrestoDB 0.267.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amt-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223,0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906,0		04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_628,0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 805.0	4,14.287	30. August 2022	us-west-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 719.0	4,14.287	10. August 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 426,0	4,14.281	10. Juni 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 406.1	4,14.275	2. Mai 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

- [Mit Amazon EMR 6.6 und höher werden Anwendungen, die Log4j 1.x und Log4j 2.x verwenden, auf Log4j 1.2.17 \(oder höher\) bzw. Log4j 2.17.1 \(oder höher\) aktualisiert, sodass die zur Behebung der Probleme bereitgestellten Bootstrap-Aktionen nicht erforderlich sind.](#) CVE
- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using](#)

[EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Sizierung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.

Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR reduziert die Cluster-Startzeit für Cluster, die die EMR AMI Standardoption verwenden und nur gängige Anwendungen wie Apache Hadoop, Apache Spark und Apache Hive installieren, um durchschnittlich bis zu 80 Sekunden.

Bekanntes Problem

- Wenn EMR Amazon-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, `NoSuchMethodError` tritt ein, weil Amazon ein falsches EMR verwendet. `Hbase.compat.version` EMR Amazon-Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. EMR Amazon-Version 6.9.0 behebt dieses Problem.
- Auf Trino-Clustern mit langer Laufzeit aktiviert Amazon EMR 6.6.0 die Garbage-Collection-Logging-Parameter in der Trino `jvm.config`, um bessere Einblicke in die Garbage-Collection-Logs zu erhalten. Durch diese Änderung werden viele Garbage-Collection-Protokolle an die Datei `launcher.log` (`/var/log/trino/launcher.log`) angehängt. Wenn Sie Trino-Cluster in Amazon EMR 6.6.0 ausführen, kann es vorkommen, dass Knoten aufgrund der angehängten Protokolle nicht mehr genügend Festplattenspeicher haben, nachdem der Cluster einige Tage lang ausgeführt wurde.

Um dieses Problem zu umgehen, führen Sie das folgende Skript als Bootstrap-Aktion aus, um die Garbage-Collection-Logging-Parameter in `jvm.config` zu deaktivieren, während Sie den Cluster für Amazon 6.6.0 erstellen oder klonen. 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"
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `RecordServer` Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. `SecretAgent` Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

6.6.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.20.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.50.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	3.2.1-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-6	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-6	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-6	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-6	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-6	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	3.2.1-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.2.1-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-6	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	3.2.1-amzn-6	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.4-amzn-2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.4-amzn-2	HBase-Befehlszeilenclient.
hbase-rest-server	2.4.4-amzn-2	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.4-amzn-2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.4-amzn-2	Reparaturtool für HBase-Apache-Cluster.
hcatalog-client	3.1.2-amzn-7	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.2-amzn-7	Bereitstellung von DienstenH Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-7	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-7	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-7	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.10.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.10.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.10.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.13.1	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.267-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	367-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	367-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	367-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.2.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.2.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.2.0-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	3.2.0-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	22.02.0-amzn-0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.

Komponente	Version	Beschreibung
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Cli-ent.

6.6.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.6.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services Resource Manager, Node Manager, Proxy Server, and Timeline Server. Additionally restarts HBase Regionserver, HBase Master, HBase Thrift, HBase Rest, Hive Server2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink <code>log4j-session.properties</code> für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der <code>raptor.properties</code> -Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der <code>redis.properties</code> -Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der <code>redshift.properties</code> -Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der <code>tpch.properties</code> -Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der <code>tpcds.properties</code> -Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei <code>dbks-site.xml</code> von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der <code>ranger-kms-site</code> XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.5.0

6.5.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
AWS SDK for 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

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.0	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.2.6	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0
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

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Presto	0,261	0,254,1	0,245,1	0,245,1
Spark	3.1.2	3.1.2	3.1.1	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	360	359	350	350
Zeppelin	0.10.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3,5.7	3.5.7	3.4.14	3.4.14

6.5.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.5.0. Änderungen beziehen sich auf Version 6.4.0.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

Neue Features

- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).
- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Sizierung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.
- Support für das offene Tabellenformat von Apache Iceberg für riesige analytische Datensätze.

- Support für ranger-trino-plugin 2.0.1-amzn-1
- Support für toree 0.5.0

Änderungen, Verbesserungen und behobene Probleme

- Die Release-Version von Amazon EMR 6.5 unterstützt jetzt Apache Iceberg 0.12.0 und bietet Laufzeitverbesserungen mit Amazon EMR Runtime for Apache Spark, Amazon EMR Runtime for Presto und Amazon EMR Runtime for Apache Hive.
- [Apache Iceberg](#) ist ein offenes Tabellenformat für große Datensätze in Amazon S3 und bietet schnelle Abfrageleistung bei großen Tabellen, atomare Commits, gleichzeitige Schreibvorgänge und SQL-kompatible Tabellenentwicklung. Mit EMR 6.5 können Sie Apache Spark 3.1.2 mit dem Iceberg-Tabellenformat verwenden.
- Apache Hudi 0.9 fügt Spark SQL DDL und Unterstützung hinzu. DML Auf diese Weise können Sie Hudi-Tabellen mit nur Anweisungen erstellen und aktualisieren. SQL Apache Hudi 0.9 beinhaltet auch Leistungsverbesserungen auf der Abfrageseite und auf der Writer-Seite.
- Amazon EMR Runtime for Apache Hive verbessert die Leistung von Apache Hive auf Amazon S3, indem Umbenennungsvorgänge bei Staging-Vorgängen entfernt werden, und verbessert die Leistung von metastore check (MSCK) -Befehlen, die zur Reparatur von Tabellen verwendet werden.

Bekanntete Probleme

- Wenn EMR Amazon-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, NoSuchMethodError tritt ein, weil Amazon ein falsches EMR verwendet. `Hbase.compat.version` EMRAmazon-Version 6.8.0 behebt dieses Problem.
- Hbase-Bundle-Cluster mit hoher Verfügbarkeit (HA) können nicht mit der Standard-Volume-Größe und dem Instance-Typ bereitgestellt werden. Dieses Problem lässt sich umgehen, indem Sie die Größe des Root-Volumens erhöhen.
- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-`workflow.xml`-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS E im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0

bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:

- Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#\$%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

6.5.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK

Komponente	Version	Beschreibung
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.4.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.19.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.48.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.14.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-5	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-5	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-5	HTTP-Endpoint für Operationen auf HDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	3.2.1-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-5	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-5	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.4.4-amzn-1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.4-amzn-1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.4.4-amzn-1	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	2.4.4-amzn-1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.4-amzn-1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-6	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-6	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-6	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-6	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-6	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-6	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.9.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.9.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.9.0-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.9.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.12.0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.261-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.261-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.261-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
trino-coordinator	360	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	360	Service für das Ausführen von Teilen einer Abfrage.
trino-client	360	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.2-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.2-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.2-amzn-1	In-Memory-Ausführungsengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	3.1.2-amzn-1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Client.

6.5.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.5.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.4.0

6.4.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
AWS SDK for 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	-	-	-	-
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

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
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
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
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

6.4.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.4.0. Änderungen beziehen sich auf Version 6.3.0.

Erste Version: 20. September 2021

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

Unterstützte Anwendungen

- AWS SDK for Java Version 1.12.31
- CloudWatch Version 2.2.0 senken
- Version 4.16.0 von EMR DynamoDB Connector
- EMRFSAusführung 2.47.0
- Amazon EMR Goodies versie 3.2.0
- Amazon EMR Kinesis Connector, Version 3.5.0
- Amazon EMR Record Server, Version 2.1.0
- Amazon EMR Scripts versie 2.5.0
- Flink, Version 1.13.1
- Ganglia Version 3.7.2
- AWS Glue Hive Metastore Client Version 3.3.0
- Hadoop Version 3.2.1-amzn-4
- HBaseAusführung 2.4.4-amzn-0
- HBase-Operator-Tools 1.1.0
- HCatalogAusführung 3.1.2-amzn-5

- Hive Version 3.1.2-amzn-5
- Hudi Version 0.8.0-amzn-0
- Hue Version 4.9.0
- JDKJava-Version Corretto-8.302.08.1 (Build 1.8.0_302-b08)
- JupyterHub Ausführung 1.4.1
- Livy Version 0.7.1-incubating
- MXNetAusführung 1.8.0
- Oozie-Version 5.2.1
- Phoenix Version 5.1.2
- Pig Version 0.17.0
- Presto Version 0.254.1-amzn-0
- Trino Version 359
- Apache Ranger KMS (transparente Multimaster-Verschlüsselung) Version 2.0.0
- Ranger-Plugins 2.0.1-amzn-0
- Ranger-S3-Plugin 1.2.0
- SageMaker Spark-Version 1.4.1 SDK
- Scala Version 2.12.10 (Offene JDK 64-Bit-Server-VM, Java 1.8.0_282)
- Spark Version 3.1.2-amzn-0
- Spark-Rapids 0.4.1
- Sqoop-Version 1.4.7
- TensorFlow Ausführung 2.4.1
- Tez-Version 0.9.2
- Zeppelin-Version 0.9.0
- Zookeeper Version 3.5.7
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

Neue Features

- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using](#)

[EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).

- Auf Apache Ranger-fähigen EMR Amazon-Clustern können Sie Apache Spark verwenden, SQL um Daten mit, und in die Apache Hive-Metastore-Tabellen einzufügen oder diese zu aktualisieren. INSERT INTO INSERT OVERWRITE ALTER TABLE Bei der Verwendung ALTER TABLE mit Spark SQL muss ein Partitionsspeicherort das untergeordnete Verzeichnis eines Tabellenspeicherorts sein. Amazon unterstützt derzeit EMR nicht das Einfügen von Daten in eine Partition, bei der sich der Speicherort der Partition vom Speicherort der Tabelle unterscheidet.
- Presto SQL wurde [in Trino umbenannt](#).
- Hive: Die Ausführung einfacher SELECT Abfragen mit LIMIT Klausel wird beschleunigt, indem die Abfrageausführung gestoppt wird, sobald die in der LIMIT Klausel angegebene Anzahl von Datensätzen abgerufen wurde. Einfache SELECT Abfragen sind Abfragen ohne ORDER BY/BY-Klausel oder Abfragen, die nicht über eine Reducer-Stufe verfügen. GROUP z. B. SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>.

Hudi-Parallelitätskontrolle

- Hudi unterstützt jetzt Optimistic Concurrency Control (OCC), das bei Schreiboperationen genutzt werden kann, z. B. UPSERT INSERT um Änderungen von mehreren Autoren an derselben Hudi-Tabelle zu ermöglichen. Dies ist auf DateiebeneOCC, sodass zwei beliebige Commits (oder Writer) in dieselbe Tabelle schreiben können, sofern ihre Änderungen nicht miteinander in Konflikt stehen. Weitere Informationen finden Sie unter [Hudi-Parallelitätskontrolle](#).
- Amazon EMR Amazon-Clustern ist Zookeeper installiert, der als Sperranbieter für verwendet werden kann. OCC Um die Verwendung dieser Funktion zu vereinfachen, sind in EMR Amazon-Clustern die folgenden Eigenschaften vorkonfiguriert:

```
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
```

Zur Aktivierung OCC müssen Sie die folgenden Eigenschaften entweder mit ihren Hudi-Joboptionen oder auf Cluster-Ebene mithilfe der Amazon-Konfigurationen konfigurieren: EMR API

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily instead of inline with every write)
```

```
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table Name is a good option)
```

Hudi Monitoring: CloudWatch Amazon-Integration zur Berichterstattung über Hudi-Metriken

- Amazon EMR unterstützt die Veröffentlichung von Hudi Metrics auf Amazon CloudWatch. Es wird aktiviert, indem die folgenden erforderlichen Konfigurationen festgelegt werden:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Im Folgenden finden Sie optionale Hudi-Konfigurationen, die Sie ändern können:

Einstellung	Beschreibung	Wert
hoodie.metrics.cloudwatch.report.period.seconds	Häufigkeit (in Sekunden), mit der Kennzahlen an Amazon gemeldet werden CloudWatch	Der Standardwert ist 60s, was für die von Amazon angebotene Standardauflösung von einer Minute in Ordnung ist CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Präfix, das jedem Metriknamen hinzugefügt werden soll	Der Standardwert ist leer (kein Präfix)
hoodie.metrics.cloudwatch.namespace	CloudWatch Amazon-Namespace, unter dem Metriken veröffentlicht werden	Der Standardwert ist Hudi
hoodie.metrics.cloudwatch.maxDatumsPerAnfrage	Maximale Anzahl von Daten, die in einer Anfrage an Amazon enthalten sein können CloudWatch	Der Standardwert ist 20, was dem CloudWatch Amazon-Standard entspricht

Unterstützung und Verbesserungen von Amazon EMR Hudi-Konfigurationen

- Kunden können jetzt die Funktion „EMRKonfigurationen API und Neukonfiguration“ nutzen, um Hudi-Konfigurationen auf Cluster-Ebene zu konfigurieren. Eine neue dateibasierte

Konfigurationsunterstützung wurde über `/etc/hudi/conf/hudi-defaults.conf` eingeführt, ähnlich wie bei anderen Anwendungen wie Spark, Hive usw. EMR konfiguriert einige Standardeinstellungen, um die Benutzererfahrung zu verbessern:

- `hoodie.datasource.hive_sync.jdbcurl` ist für den Cluster-Hive-Server konfiguriert URL und muss nicht mehr angegeben werden. Dies ist besonders nützlich, wenn Sie einen Job im Spark-Cluster-Modus ausführen, wo Sie zuvor die EMR Amazon-Master-IP angeben mussten.
- HBase spezifische Konfigurationen, die für die Verwendung des HBase Index mit Hudi nützlich sind.
- Spezifische Konfiguration des Zookeeper-Sperranbieters, wie unter Parallelitätskontrolle beschrieben, was die Verwendung von Optimistic Concurrency Control () erleichtert. OCC
- Zusätzliche Änderungen wurden eingeführt, um die Anzahl der Konfigurationen zu reduzieren, die Sie bestehen müssen, und um nach Möglichkeit automatische Schlüsse zu ziehen:
 - Das `partitionBy` -Schlüsselwort kann verwendet werden, um die Partitionsspalte zu spezifizieren.
 - Bei der Aktivierung von Hive Sync ist es nicht mehr erforderlich, den Vorgang `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY` zu bestehen. Diese Werte können aus dem Hudi-Tabellennamen und dem Partitionsfeld abgeleitet werden.
 - `KEYGENERATOR_CLASS_OPT_KEY` ist nicht zwingend erforderlich und kann aus einfacheren Fällen von `SimpleKeyGenerator` und `ComplexKeyGenerator` abgeleitet werden.

Vorbehalte von Hudi

- Hudi unterstützt keine vektorisierte Ausführung in Hive für Merge on Read (MoR)- und Bootstrap-Tabellen. So schlägt beispielsweise `count(*)` mit der Hudi-Echtzeittabelle fehl, wenn `hive.vectorized.execution.enabled` auf „wahr“ gesetzt ist. Um das Problem zu umgehen, können Sie das vektorisierte Lesen deaktivieren, indem Sie `hive.vectorized.execution.enabled` auf `false` setzen.
- Die Multi-Writer-Unterstützung ist nicht mit dem Hudi-Bootstrap-Feature kompatibel.
- Flink Streamer und Flink SQL sind experimentelle Funktionen in dieser Version. Diese Features werden nicht zur Verwendung in Produktionsbereitstellungen empfohlen.

Änderungen, Verbesserungen und behobene Probleme

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

- Bisher führte ein manueller Neustart des Resource Managers auf einem Multi-Master-Cluster dazu, dass EMR Amazon-On-Cluster-Daemons wie Zookeeper alle zuvor außer Betrieb genommenen oder verloren gegangenen Knoten in der Zookeeper-Znode-Datei neu laden. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon entfernt EMR jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei und die internen Grenzwerte wurden erhöht.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Konfiguration eines Clusters zur Behebung von Leistungsproblemen mit Apache YARN Timeline Server Version 1 und 1.5

Apache YARN Timeline Server Version 1 und 1.5 können bei sehr aktiven, großen EMR Clustern zu Leistungsproblemen führen, insbesondere bei `yarn.resourcemanager.system-metrics-publisher.enabled=true`, was die Standardeinstellung in Amazon ist EMR. Ein YARN Open-Source-Timeline Server v2 löst das Leistungsproblem im Zusammenhang mit der Skalierbarkeit von YARN Timeline Server.

Weitere Lösungen für dieses Problem umfassen:

- Konfiguration von `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- Aktivieren des Fixes für dieses Problem beim Erstellen eines Clusters, wie unten beschrieben.

Die folgenden EMR Amazon-Versionen enthalten eine Lösung für dieses Leistungsproblem mit dem YARN 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

Um den Fix für eine der oben angegebenen EMR Amazon-Versionen zu aktivieren, legen Sie diese Eigenschaften `true` in einer JSON Konfigurationsdatei fest, die mit dem [aws emr create-cluster](#) Befehlsparameter übergeben wird: `--configurations file://./configurations.json`. Oder aktivieren Sie den Fix über die Benutzeroberfläche der [Rekonfigurationskonsole](#).

Beispiel für den Inhalt der Datei `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": []
}
```

- Web HDFS - und HttpFS-Server sind standardmäßig deaktiviert. Sie können Web HDFS mithilfe der Hadoop-Konfiguration erneut aktivieren. `dfs.webhdfs.enabled` Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPCE S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys](#).
- Hive: Die Leistung bei Schreibabfragen wurde verbessert, indem die Verwendung eines Scratch-Verzeichnisses HDFS für den letzten Job aktiviert wurde. Die temporären Daten für den endgültigen Job werden HDFS nicht in Amazon S3 geschrieben, und die Leistung wird verbessert, da die Daten vom HDFS Speicherort der endgültigen Tabelle (Amazon S3) und nicht zwischen Amazon S3 S3-Geräten verschoben werden.
- Hive: Verbesserung der Kompilierungszeit von Abfragen um das 2,5-fache mit Glue Metastore Partition Pruning.
- Wenn Built-In von Hive an den Hive Metastore Server übergeben UDFs werden, wird standardmäßig nur ein Teil der UDFs Built-In an den Glue Metastore übergeben, da Glue nur begrenzte Ausdrucksoperatoren unterstützt. Wenn Sie `hive.glue.partition.pruning.client=true` festlegen, erfolgt das gesamte Partitionsbereinigen auf der Clientseite. Wenn Sie `hive.glue.partition.pruning.server=true` festlegen, erfolgt das gesamte Bereinigen von Partitionen auf der Serverseite.

Bekannte Probleme

- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache Hadoop HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem EMR primären Amazon-Node entweder manuell mithilfe `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen](#) Amazon-Schritt. EMR
- Die Amazon EMR Notebooks-Funktion, die mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den der Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das Notebook mit dem EMR Cluster verbinden. `sudo systemctl start hadoop-httpfs`
- In EMR Amazon-Version 6.4.0 unterstützt Phoenix die Komponente Phoenix Connectors nicht.

- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-workflow.xml-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS E im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist s3://bucket/table/p=a ein Präfix von s3://bucket/table/p=a b.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in s3://bucket/table/p=a b zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#\$%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

6.4.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.3.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.47.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.13.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
flink-jobmanager-config	1.13.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-https-server	3.2.1-amzn-4	HTTP-Endpoint für Operationen in HDFS.
hadoop-kms-server	3.2.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.2.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-4	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-0	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.4.4-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.4-amzn-0	HBase-Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	2.4.4-amzn-0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.4.4-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-5	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-5	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-5	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-5	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-5	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-5	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.8.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.8.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.8.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.8.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit

Komponente	Version	Beschreibung
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.1.2	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.254.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.254.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.254.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	359	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.

Komponente	Version	Beschreibung
trino-worker	359	Service für das Ausführen von Teilen einer Abfrage.
trino-client	359	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4,0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.2-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.2-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.2-amzn-0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	3.1.2-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Client.

6.4.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die

Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.4.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode,

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBase Regionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.3.1

6.3.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
AWS SDK for 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	-	-	-	-
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

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
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
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
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

6.3.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket

nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.

- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPC S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys.](#)

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

6.3.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.46.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-3.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-3.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-3.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-3.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-3.1	HTTP-Endpoint für Operationen mit HDFS.
hadoop-kms-server	3.2.1-amzn-3.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.2.1-amzn-3.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-3.1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	2.2.6-amzn-1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	2.2.6-amzn-1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.6-amzn-1	HBase Befehlszeilenclient.
hbase-rest-server	2.2.6-amzn-1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.6-amzn-1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-4	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	3.1.2-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.7.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.7.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.7.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi. SQL
hudi-spark	0.7.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0- -2.0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
prestoql-coordinator	350	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	350	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	350	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.1-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.1-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.1-amzn-0.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	3.1.1-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

6.3.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.3.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in der Datei SQL node.properties von Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-blackhole	Ändern Sie die Werte in der Datei blackhole.properties von PrestoSQL.	Not available.
prestosql-connector-cassandra	Ändern Sie die Werte in der Datei cassandra.properties von PrestoSQL.	Not available.
prestosql-connector-hive	Ändern Sie die Werte in der Datei hive.properties von Presto. SQL	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Ändern Sie die Werte in der Datei jmx.properties von PrestoSQL.	Not available.
prestosql-connector-kafka	Ändern Sie die Werte in der Datei kafka.properties von PrestoSQL.	Not available.
prestosql-connector-localfile	Ändern Sie die Werte in der Datei localfile.properties von PrestoSQL.	Not available.
prestosql-connector-memory	Ändern Sie die Werte in der Datei memory.properties von PrestoSQL.	Not available.
prestosql-connector-mongodb	Ändern Sie die Werte in der Datei mongodb.properties von PrestoSQL.	Not available.
prestosql-connector-mysql	Ändern Sie die Werte in der Datei mysql.properties von PrestoSQL.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-postgresql	Ändern Sie die Werte in der Datei postgresql.properties von PrestoSQL.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in der Datei raptor.properties von Presto. SQL	Not available.
prestosql-connector-redis	Ändern Sie die Werte in der Datei redis.properties von PrestoSQL.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in der Datei redshift.properties von PrestoSQL.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in der Datei tpch.properties von Presto. SQL	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in der Datei tpcds.properties von PrestoSQL.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.3.0

6.3.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
AWS SDK for 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

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
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
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

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
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

6.3.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.3.0. Änderungen beziehen sich auf Version 6.2.0.

Erste Version: 12. Mai 2021

Letzte Aktualisierung: 9. August 2021

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.977
- CloudWatch Version 2.1.0 senken
- Version 4.16.0 von EMR DynamoDB Connector
- EMRFSAusführung 2.46.0
- Amazon EMR Goodies versie 3.2.0
- Amazon EMR Kinesis Connector, Version 3.5.0
- Amazon EMR Record Server, Version 2.0.0
- Amazon EMR Scripts versie 2.5.0
- Flink-Version 1.12.1
- Ganglia Version 3.7.2
- AWS Glue Hive Metastore Client Version 3.2.0
- Hadoop-Version 3.2.1-amzn-3
- HBaseAusführung 2.2.6-amzn-1
- HBase-Operator-Tools 1.0.0

- HCatalogAusführung 3.1.2-amzn-0
- Hive-Version 3.1.2-amzn-4
- Hudi-Version 0.7.0-amzn-0
- Hue Version 4.9.0
- JDKJava-Version Corretto-8.282.08.1 (Build 1.8.0_282-b08)
- JupyterHub Ausführung 1.2.0
- Livy-Version 0.7.0-incubating
- MXNetAusführung 1.7.0
- Oozie-Version 5.2.1
- Phoenix Version 5.0.0
- Pig Version 0.17.0
- Presto-Version 0.245.1-amzn-0
- Presto, Version 3.5.0 SQL
- Apache Ranger KMS (transparente Multimaster-Verschlüsselung) Version 2.0.0
- Ranger-Plugins 2.0.1-amzn-0
- Ranger-S3-Plugin 1.1.0
- SageMaker Spark-Version 1.4.1 SDK
- Scala Version 2.12.10 (Offene JDK 64-Bit-Server-VM, Java 1.8.0_282)
- Spark-Version 3.1.1-amzn-0
- Spark-Rapids 0.4.1
- Sqoop-Version 1.4.7
- TensorFlow Ausführung 2.4.1
- Tez-Version 0.9.2
- Zeppelin-Version 0.9.0
- Zookeeper Version 3.4.14
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

Neue Features

- Amazon EMR unterstützt Amazon S3 Access Points, eine Funktion von Amazon S3, mit der Sie den Zugriff auf gemeinsam genutzte Data Lakes einfach verwalten können. Mit Ihrem

Amazon S3 Access Point-Alias können Sie Ihren Datenzugriff auf Amazon in großem Umfang vereinfachen. Sie können Amazon S3 Access Points mit allen Versionen von Amazon EMR ohne zusätzliche Kosten in allen AWS Regionen verwenden, in denen Amazon verfügbar ist. Weitere Informationen zu Amazon S3 Access Points und Zugangspunkt-Aliassen finden Sie unter [Verwenden eines Alias im Bucket-Stil für Ihren Zugangspunkt](#) im Amazon-S3-Benutzerhandbuch.

- Die neuen `ListReleaseLabel` API Parameter `DescribeReleaseLabel` und die Parameter enthalten Informationen zum EMR Amazon-Release-Label. Sie können programmgesteuert Versionen auflisten, die in der Region verfügbar sind, in der die API Anfrage ausgeführt wird, und die verfügbaren Anwendungen für ein bestimmtes EMR Amazon-Release-Label auflisten. In den Release-Label-Parametern werden auch EMR Amazon-Releases aufgeführt, die eine bestimmte Anwendung unterstützen, z. B. Spark. Diese Informationen können verwendet werden, um EMR Amazon-Cluster programmgesteuert zu starten. Sie können beispielsweise einen Cluster mit der neuesten Release-Version aus den `ListReleaseLabel`-Ergebnissen starten. Weitere Informationen finden Sie unter [DescribeReleaseLabel](#) und [ListReleaseLabels](#) in der EMR API Amazon-Referenz.
- Mit Amazon EMR 6.3.0 können Sie einen Cluster starten, der nativ in Apache Ranger integriert ist. Apache Ranger ist ein Open-Source-Framework zur Aktivierung, Überwachung und Verwaltung einer umfassenden Datensicherheit auf der gesamten Hadoop-Plattform. Weitere Informationen finden Sie unter [Apache Ranger](#). Dank der nativen Integration können Sie Ihren eigenen Apache Ranger verwenden, um eine detaillierte Datenzugriffskontrolle auf Amazon durchzusetzen. EMR Weitere Informationen finden Sie unter [Integrieren von Amazon EMR mit Apache Ranger](#) im Amazon EMR Management Guide.
- **Verwaltete Richtlinien mit Geltungsbereich:** Um den AWS bewährten Verfahren Rechnung zu tragen, hat Amazon EMR verwaltete Standardrichtlinien mit Geltungsbereich v2 als Ersatz für Richtlinien eingeführt, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [EMR Von Amazon verwaltete Richtlinien](#).
- **Support-Status für Instance Metadata Service (IMDS) V2:** Für Amazon EMR 6.2 oder höher werden EMR Amazon-Komponenten IMDSv2 für alle IMDS Aufrufe verwendet. Für IMDS Aufrufe in Ihrem Anwendungscode können Sie IMDSv1 sowohl als auch verwenden oder die Option so konfigurieren `IMDSv2`, IMDS dass sie nur IMDSv2 für zusätzliche Sicherheit verwendet wird. Wenn Sie die Option `IMDSv1` in früheren Versionen von Amazon EMR 6.x deaktivieren, führt dies zu einem Fehler beim Starten des Clusters.

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Die SQL Standardeinstellung für den `formatted` Erklärmodus [der](#) Spark-Benutzeroberfläche wurde in Spark 3.1 von `extended` auf `formatted` geändert. Amazon EMR hat es wieder auf die Aufnahme `extended` logischer Planinformationen in die SQL Spark-Benutzeroberfläche zurückgesetzt. Dies

kann durch Einstellen von `spark.sql.ui.explainMode` auf `formatted` rückgängig gemacht werden.

- Die folgenden Commits wurden aus dem Spark-Master-Branch zurückportiert.
 - [\[SPARK-34752\]](#) [\[BUILD\]](#) Geben Sie Jetty an 9.4.37 an die Adresse -2020-27223. CVE
 - [\[-34534\]](#) [Korrigiert die Reihenfolge beim Abrufen von Blöcken. SPARK](#) `blockIds`
`FetchShuffleBlocks`
 - [\[SPARK-34681\]](#) [\[SQL\]](#) Es wurde ein Fehler behoben, der beim Erstellen der linken Seite mit einer ungleichen Bedingung beim Erstellen eines kompletten Router-Shuffled-Hash-Joins auftrat.
 - [\[SPARK-34497\]](#) [\[SQL\]](#) Integrierte Verbindungsanbieter wurden behoben, um Änderungen im Sicherheitskontext wiederherzustellen. JDBC JVM
- [Um die Interoperabilität mit dem Nvidia RAPIDs Spark-Plugin zu verbessern, wurde eine Problemumgehung hinzugefügt, um ein Problem zu beheben, das verhindert, dass dynamisches Partitionsbereinigen ausgelöst wird, wenn Nvidia Spark RAPIDs mit deaktivierter adaptiver Abfrageausführung verwendet wird. Weitere Informationen finden Sie unter RAPIDSPROBLEM #1378 und Problem # #1386. RAPIDS](#) Einzelheiten zur neuen Konfiguration `spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse` finden Sie in [RAPIDSAusgabe #1386](#).
- Der Standardalgorithmus für den Dateiausgabe-Committer wurde in Open Source Spark 3.1 vom v2-Algorithmus auf den v1-Algorithmus geändert. Weitere Informationen finden Sie unter [Amazon EMR Optimizing Spark Performance — Dynamic Partition Pruning](#).
- Amazon EMR kehrte zum v2-Algorithmus zurück, dem in früheren Amazon EMR 6.x-Versionen verwendetem Standard, um Leistungsrückgänge zu verhindern. Um das Open-Source-Verhalten von Spark 3.1 wiederherzustellen, setzen Sie `spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version` auf 1. Open Source Spark hat diese Änderung vorgenommen, weil das Festschreiben von Aufgaben im Dateioutput-Committer-Algorithmus v2 nicht atomar ist, was in einigen Fällen zu Problemen mit der Korrektheit der Ausgabedaten führen kann. Das Festschreiben von Aufgaben im Algorithmus v1 ist jedoch auch nicht atomar. In einigen Szenarien beinhaltet die Aufgabenübergabe ein Löschen, das vor einer Umbenennung durchgeführt wird. Dies kann zu einem unbemerkten Problem mit der Datenkorrektheit führen.
- Managed Scaling-Probleme in früheren EMR Amazon-Versionen wurden behoben und Verbesserungen vorgenommen, sodass die Ausfallraten von Anwendungen erheblich reduziert wurden.

- Das AWS SDK Java-Bundle wurde auf jedem neuen Cluster installiert. Dies ist ein einzelnes JAR, das alle Dienste SDKs und ihre Abhängigkeiten enthält, anstatt einzelne Komponenten-Jars. Weitere Informationen finden Sie unter [Java SDK Bundled Dependency](#).

Bekannte Probleme

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Websites wie SparkUIs, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- Wenn der AWS Glue-Datenkatalog aktiviert ist, URI kann die Verwendung von Spark für den Zugriff auf eine AWS Glue-Datenbank mit einem Null-String-Speicherort fehlschlagen. Dies passiert bei früheren EMR Amazon-Versionen, aber mit SPARK -31709 (<https://issues.apache.org/jira/browse/SPARK-31709>) gilt es für mehr Fälle. Wenn Sie beispielsweise eine Tabelle in der AWS Standard-Glue-DB erstellen, deren Speicherort eine Nullzeichenfolge URI ist, `spark.sql("CREATE TABLE mytest (key string) location '/table_path';")` schlägt dies mit der Meldung fehl, dass kein Pfad aus einer leeren Zeichenfolge erstellt werden kann. Um dies zu umgehen, legen Sie manuell einen Speicherort URI für Ihre AWS Glue-Datenbanken fest und erstellen Sie dann mit Spark Tabellen innerhalb dieser Datenbanken.
- In Amazon EMR 6.3.0 SQL wurde Presto von Version 343 auf Version 350 aktualisiert. Es gibt zwei sicherheitsrelevante Änderungen von Open Source, die sich auf diese Versionsänderung beziehen. Die dateibasierte Katalogzugriffssteuerung wurde von `deny` zu `allow` geändert, wenn keine Regeln für Tabellen-, Schema- oder Sitzungseigenschaften definiert sind. Außerdem wurde die dateibasierte Systemzugriffskontrolle dahingehend geändert, dass sie Dateien unterstützt, für die keine Katalogregeln definiert sind. In diesem Fall ist der gesamte Zugriff auf Kataloge zulässig.

Weitere Informationen finden Sie in [Version 344 \(9. Oktober 2020\)](#).

- Beachten Sie, dass das Hadoop-Benutzerverzeichnis (`/home/hadoop`) für jeden lesbar ist. Es verfügt über Unix `755 (drwxr-xr-x)`-Verzeichnisberechtigungen, um Frameworks wie Hive den Lesezugriff zu ermöglichen. Sie können Dateien in `/home/hadoop` und seinen Unterverzeichnissen ablegen, beachten Sie jedoch die Berechtigungen für diese Verzeichnisse, um vertrauliche Informationen zu schützen.

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMRAmazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standardeinstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein `ulimit` über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein `ulimit` festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das `ulimit` für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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

EMRCluster, auf denen Amazon Linux oder Amazon Linux 2 Amazon Machine Images (AMIs) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen EC2 Amazon-Instances, auf denen das standardmäßige Amazon Linux ausgeführt wird. Wenn neue Amazon Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel und CUDA Updates) NVIDIA, verfügbar werden, nachdem eine EMR Amazon-Version verfügbar wird, laden EMR Cluster-Instances, die standardmäßig ausgeführt werden, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon so anpassen EMR AMI](#), dass es [das neueste Amazon Linux verwendet AMI](#).

- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-workflow.xml-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFSE im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```


- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

6.3.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMRS3Select-Anschluss
emrfs	2.46.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-3	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-3	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-3	HTTPEndpunkt für OperationenHDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	3.2.1-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-3	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.2.6-amzn-1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.2.6-amzn-1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.6-amzn-1	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	2.2.6-amzn-1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.6-amzn-1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-4	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.7.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.7.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.7.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi. SQL
hudi-spark	0.7.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0- -2.0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
prestoql-coordinator	350	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	350	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	350	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.1-amzn-0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	3.1.1-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

6.3.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.3.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in der Datei SQL node.properties von Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestoql-connector-blackhole	Ändern Sie die Werte in der Datei blackhole.properties von PrestoSQL.	Not available.
prestoql-connector-cassandra	Ändern Sie die Werte in der Datei cassandra.properties von PrestoSQL.	Not available.
prestoql-connector-hive	Ändern Sie die Werte in der Datei hive.properties von Presto. SQL	Restarts Presto-Server (for PrestoSQL)
prestoql-connector-jmx	Ändern Sie die Werte in der Datei jmx.properties von PrestoSQL.	Not available.
prestoql-connector-kafka	Ändern Sie die Werte in der Datei kafka.properties von PrestoSQL.	Not available.
prestoql-connector-localfile	Ändern Sie die Werte in der Datei localfile.properties von PrestoSQL.	Not available.
prestoql-connector-memory	Ändern Sie die Werte in der Datei memory.properties von PrestoSQL.	Not available.
prestoql-connector-mongodb	Ändern Sie die Werte in der Datei mongodb.properties von PrestoSQL.	Not available.
prestoql-connector-mysql	Ändern Sie die Werte in der Datei mysql.properties von PrestoSQL.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-postgresql	Ändern Sie die Werte in der Datei postgresql.properties von PrestoSQL.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in der Datei raptor.properties von Presto. SQL	Not available.
prestosql-connector-redis	Ändern Sie die Werte in der Datei redis.properties von PrestoSQL.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in der Datei redshift.properties von PrestoSQL.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in der Datei tpch.properties von Presto. SQL	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in der Datei tpcds.properties von PrestoSQL.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.2.1

6.2.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
AWS SDK for 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

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
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
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

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
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

6.2.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren,

fehlschlugen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.

- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPC S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys.](#)

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

6.2.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.0.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
<code>emr-s3-dist-cp</code>	2.16.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	2.0.0	EMRS3Select-Anschluss
emrfs	2.44.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-2.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-2.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.2.1-amzn-2.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-2.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-https-server	3.2.1-amzn-2.1	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	3.2.1-amzn-2.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.2.1-amzn-2.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-2.1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.2.6-amzn-0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.2.6-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.6-amzn-0	HBase Befehlszeilenclient.
hbase-rest-server	2.2.6-amzn-0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.6-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-3	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	3.1.2-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-3	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	3.1.2-amzn-3	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.6.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.6.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.6.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi. SQL
hudi-spark	0.6.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5,0.0- -2,0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.238.3-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.238.3-amzn-1	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.238.3-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
prestoql-coordinator	343	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	343	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	343	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.1-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.1-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	3.0.1-amzn-0.1	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	3.0.1-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	0.2.0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

6.2.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.2.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Not available.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Not available.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Not available.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in der Datei SQL node.properties von Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestoql-connector-blackhole	Ändern Sie die Werte in der Datei blackhole.properties von PrestoSQL.	Not available.
prestoql-connector-cassandra	Ändern Sie die Werte in der Datei cassandra.properties von PrestoSQL.	Not available.
prestoql-connector-hive	Ändern Sie die Werte in der Datei hive.properties von Presto. SQL	Restarts Presto-Server (for PrestoSQL)
prestoql-connector-jmx	Ändern Sie die Werte in der Datei jmx.properties von PrestoSQL.	Not available.
prestoql-connector-kafka	Ändern Sie die Werte in der Datei kafka.properties von PrestoSQL.	Not available.
prestoql-connector-localfile	Ändern Sie die Werte in der Datei localfile.properties von PrestoSQL.	Not available.
prestoql-connector-memory	Ändern Sie die Werte in der Datei memory.properties von PrestoSQL.	Not available.
prestoql-connector-mongodb	Ändern Sie die Werte in der Datei mongodb.properties von PrestoSQL.	Not available.
prestoql-connector-mysql	Ändern Sie die Werte in der Datei mysql.properties von PrestoSQL.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-postgresql	Ändern Sie die Werte in der Datei postgresql.properties von PrestoSQL.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in der Datei raptor.properties von Presto. SQL	Not available.
prestosql-connector-redis	Ändern Sie die Werte in der Datei redis.properties von PrestoSQL.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in der Datei redshift.properties von PrestoSQL.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in der Datei tpch.properties von Presto. SQL	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in der Datei tpcds.properties von PrestoSQL.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.2.0

6.2.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
AWS SDK for Java	1.11.880	1.11.828	1.11.828	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	2.12.10
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

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
HBase	2.2.6-amzn-0	2.2.5	2.2.5	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1
Hue	4.8.0	4.7.1	4.7.1	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	-
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

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
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

6.2.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.2.0. Änderungen beziehen sich auf Version 6.1.0.

Veröffentlichungsdatum: 9. Dezember 2020

Letzte Aktualisierung: 4. Oktober 2021

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.828
- emr-record-server Ausführung 1.7.0
- Flink-Version 1.11.2
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBaseAusführung 2.2.6-amzn-0
- HBase-Operator-Tools 1.0.0
- HCatalogAusführung 3.1.2-amzn-0
- Hive Version 3.1.2-amzn-3
- Hudi Version 0.6.0-amzn-1
- Hue Version 4.8.0
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0

- MXNetAusführung 1.7.0
- Oozie-Version 5.2.0
- Phoenix Version 5.0.0
- Pig Version 0.17.0
- Presto Version 0.238.3-amzn-1
- Presto, Version 3.4.3 SQL
- Spark Version 3.0.1-amzn-0
- Spark-Rapids 0.2.0
- TensorFlow Ausführung 2.3.1
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

Neue Features

- HBase: Die Umbenennung in der Commit-Phase wurde entfernt und persistentes HFile Tracking hinzugefügt. Weitere Informationen finden Sie unter [Persistent HFile Tracking](#) im Amazon EMR Release Guide.
- HBase: Backported [Erstellen Sie eine Konfiguration, die das Zwischenspeichern von Blöcken bei der Komprimierung erzwingt](#).
- PrestoDB: Verbesserungen beim dynamischen Partitionsbereinigen. Die regelbasierte Join Reorder funktioniert mit nicht partitionierten Daten.
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Verfahren Rechnung zu tragen, EMR hat Amazon EMR verwaltete Standardrichtlinien mit Geltungsbereich v2 als Ersatz für Richtlinien eingeführt, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [EMRVon Amazon verwaltete Richtlinien](#).
- Support-Status für Instance Metadata Service (IMDS) V2: Für Amazon EMR 6.2 oder höher werden EMR Amazon-Komponenten IMDSv2 für alle IMDS Aufrufe verwendet. Für IMDS Aufrufe in Ihrem Anwendungscode können Sie IMDSv1 sowohl als auch verwenden oder die Option so konfigurierenIMDSv2, IMDS dass sie nur IMDSv2 für zusätzliche Sicherheit verwendet wird. Wenn Sie die Option IMDSv1 in früheren Versionen von Amazon EMR 6.x deaktivieren, führt dies zu einem Fehler beim Starten des Clusters.

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Spark: Leistungsverbesserungen in der Spark-Laufzeit.

Bekannte Probleme

- Amazon EMR 6.2 hat falsche Berechtigungen für die Datei `libinstance-controller-java /etc/cron.d/` in 6.2.0 festgelegt. EMR Die Berechtigungen für die Datei sind 645 (-rw-r--r-x), obwohl sie 644 (-rw-r--r--) sein sollten. Aus diesem Grund protokolliert Amazon EMR Version 6.2 keine Instance-State-Logs und das Verzeichnis `/emr/instance-logs` ist leer. Dieses Problem wurde in Amazon EMR 6.3.0 und höher behoben.

Zur Umgehung dieses Problems führen Sie das folgende Skript als Bootstrap-Aktion beim Clusterstart aus.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Für private Subnetz-Cluster von Amazon EMR 6.2.0 und 6.3.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Websites wie SparkUIs, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den `httpd`-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- In Amazon EMR 6.2.0 gibt es ein Problem, bei dem `httpd` kontinuierlich ausfällt, wodurch Ganglia nicht verfügbar ist. Sie erhalten die Fehlermeldung „Es kann keine Verbindung zum Server hergestellt werden“. Um einen Cluster zu reparieren, der bereits mit diesem Problem läuft, gehen Sie SSH zum primären Clusterknoten und fügen Sie die Zeile `Listen 80` zu der Datei `httpd.conf` hinzu, die sich unter befindet. `/etc/httpd/conf/httpd.conf` Dieses Problem wurde in Amazon EMR 6.3.0 behoben.
- HTTPDSchlägt auf EMR 6.2.0-Clustern fehl, wenn Sie eine Sicherheitskonfiguration verwenden. Dadurch ist die Benutzeroberfläche der Ganglia-Webanwendung nicht verfügbar. Um auf die Benutzeroberfläche der Ganglia-Webanwendung zuzugreifen, fügen Sie `Listen 80` der `/etc/httpd/conf/httpd.conf`-Datei auf dem Primärknoten Ihres Clusters etwas hinzu. Informationen zum Herstellen einer Verbindung mit Ihrem Cluster finden Sie unter [Herstellen einer Verbindung zum Primärknoten mithilfe von SSH](#).

EMRNotebooks können auch keine Verbindung zu EMR 6.2.0-Clustern herstellen, wenn Sie eine Sicherheitskonfiguration verwenden. Das Notebook kann keine Kernel auflisten und Spark-Aufträge

nicht weiterleiten. Wir empfehlen, EMR Notebooks EMR stattdessen mit einer anderen Version von Amazon zu verwenden.

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein `ulimit` über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein `ulimit` festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich entscheidend auf alle Hudi-Einfüge-, Upsert- und Löschvorgänge auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi zu erhalten. RPM

• **⚠ Important**

EMRCluster, auf denen Amazon Linux oder Amazon Linux 2 Amazon Machine Images (AMIs) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen EC2 Amazon-Instances, auf denen das standardmäßige Amazon Linux ausgeführt wird. AMI. Wenn neue Amazon Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel und CUDA Updates) NVIDIA, verfügbar werden, nachdem eine EMR Amazon-Version verfügbar wird, laden EMR Cluster-Instances, die standardmäßig ausgeführt werden, diese Updates AMI

nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon so anpassen EMR AMI](#), dass es [das neueste Amazon Linux verwendet AMI](#).

- Amazon EMR 6.2.0 Maven-Artefakte werden nicht veröffentlicht. Sie werden mit einer future Version von Amazon veröffentlichtEMR.
- Die persistente HFile Nachverfolgung mithilfe der HBase Storefile-Systemtabelle unterstützt die Funktion zur HBase Regionsreplikation nicht. Weitere Informationen zur HBase Regionsreplikation finden Sie unter [Timeline-consistent](#) High Available Reads.
- Unterschiede zwischen den Bucketing-Versionen von Amazon EMR EMR 6.x und 5.x Hive

EMR5.x verwendet OOS Apache Hive 2, während in EMR 6.x Apache Hive 3 verwendet wird. OOS Die Open-Source-Version Hive2 verwendet Bucketing Version 1, während die Open-Source-Version Hive3 Bucketing Version 2 verwendet. Dieser Unterschied in der Bucketing-Version zwischen Hive 2 (EMR5.x) und Hive 3 (EMR6.x) bedeutet, dass Hive Bucketing-Hashing anders funktioniert. Sehen Sie sich das folgende Beispiel an.

Die folgende Tabelle ist ein Beispiel, das in 6.x bzw. 5.x erstellt wurde. EMR EMR

```
-- 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/';
```

Dieselben Daten werden sowohl in EMR 6.x als auch in 5.x eingefügt. EMR

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Die Überprüfung des S3-Speicherorts zeigt, dass der Name der Bucketing-Datei unterschiedlich ist, da sich die Hashing-Funktion zwischen EMR 6.x (Hive 3) und 5.x (Hive 2) unterscheidet. EMR

```
[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
```

Sie können den Versionsunterschied auch erkennen, indem Sie den folgenden Befehl in Hive in 6.x ausführen. CLI EMR Beachten Sie, dass die Bucketing-Version 2 zurückgegeben wird.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

6.2.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.0.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.16.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.0.0	EMRS3Select-Anschluss
emrfs	2.44.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-2	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-2	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-2	HTTP-Endpoint für Operationen in HDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	3.2.1-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-2	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.2.6-amzn-0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.2.6-amzn-0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.6-amzn-0	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	2.2.6-amzn-0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.6-amzn-0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-3	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-3	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-3	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.6.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.6.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi. SQL
hudi-spark	0.6.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5,0.0- -2,0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.238.3-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.238.3-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
prestoql-coordinator	343	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	343	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	343	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.0.1-amzn-0	In-Memory-Ausführungseengine fürYARN.

Komponente	Version	Beschreibung
spark-yarn-slave	3.0.1-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
spark-rapids	0.2.0	Nvidia RAPIDS Spark-Plugin, das Apache Spark beschleunigt mit GPUs.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

6.2.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-6.2.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> von Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Not available.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Not available.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Werte in der HDFS Umgebung ändern.	Restarts Hadoop HDFS ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>hive-exec-log</code>	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>	Not available.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.	Not available.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
presto-sql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)
presto-sql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-node	Ändern Sie die Werte in der Datei SQL node.properties von Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-blackhole	Ändern Sie die Werte in der Datei blackhole.properties von PrestoSQL.	Not available.
prestosql-connector-cassandra	Ändern Sie die Werte in der Datei cassandra.properties von PrestoSQL.	Not available.
prestosql-connector-hive	Ändern Sie die Werte in der Datei hive.properties von Presto. SQL	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Ändern Sie die Werte in der Datei jmx.properties von PrestoSQL.	Not available.
prestosql-connector-kafka	Ändern Sie die Werte in der Datei kafka.properties von PrestoSQL.	Not available.
prestosql-connector-localfile	Ändern Sie die Werte in der Datei localfile.properties von PrestoSQL.	Not available.
prestosql-connector-memory	Ändern Sie die Werte in der Datei memory.properties von PrestoSQL.	Not available.
prestosql-connector-mongodb	Ändern Sie die Werte in der Datei mongodb.properties von PrestoSQL.	Not available.
prestosql-connector-mysql	Ändern Sie die Werte in der Datei mysql.properties von PrestoSQL.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-postgresql	Ändern Sie die Werte in der Datei postgresql.properties von PrestoSQL.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in der Datei raptor.properties von Presto. SQL	Not available.
prestosql-connector-redis	Ändern Sie die Werte in der Datei redis.properties von PrestoSQL.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in der Datei redshift.properties von PrestoSQL.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in der Datei tpch.properties von Presto. SQL	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in der Datei tpcds.properties von PrestoSQL.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 6.1.1

6.1.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for 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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

6.1.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren,

fehlschlugen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.

- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPC S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys.](#)

6.1.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK

Komponente	Version	Beschreibung
emr-ddb	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.0.0	EMRS3Select-Anschluss
emrfs	2.42.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-1.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-1.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-1.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-1.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-1.1	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	3.2.1-amzn-1.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	3.2.1-amzn-1.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-1.1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.2.5	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.2.5	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.5	HBase Befehlszeilenclient.
hbase-rest-server	2.2.5	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.5	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.2-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-2	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating-amzn-2	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi. SQL

Komponente	Version	Beschreibung
hudi-spark	0.5.2-incubating-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0- -2.0 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
prestoql-coordinator	338	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	338	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	338	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.0-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.0-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.0.0-amzn-0.1	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	3.0.0-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

6.1.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-6.1.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> YARN von Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.

Klassifizierungen	Beschreibung
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-env	Werte in der HDFS Umgebung ändern.
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-props-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. hive-exec-log
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.

Klassifizierungen	Beschreibung
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
prestosql-node	Ändern Sie die Werte in der Datei SQL node.properties von Presto.
prestosql-connector-blackhole	Ändern Sie die Werte in der Datei blackhole.properties von PrestoSQL.
prestosql-connector-cassandra	Ändern Sie die Werte in der Datei cassandra.properties von PrestoSQL.
prestosql-connector-hive	Ändern Sie die Werte in der Datei hive.properties von Presto. SQL
prestosql-connector-jmx	Ändern Sie die Werte in der Datei jmx.properties von PrestoSQL.
prestosql-connector-kafka	Ändern Sie die Werte in der Datei kafka.properties von PrestoSQL.
prestosql-connector-localfile	Ändern Sie die Werte in der Datei localfile.properties von PrestoSQL.
prestosql-connector-memory	Ändern Sie die Werte in der Datei memory.properties von PrestoSQL.
prestosql-connector-mongodb	Ändern Sie die Werte in der Datei mongodb.properties von PrestoSQL.
prestosql-connector-mysql	Ändern Sie die Werte in der Datei mysql.properties von PrestoSQL.
prestosql-connector-postgresql	Ändern Sie die Werte in der Datei postgresql.properties von PrestoSQL.

Klassifizierungen	Beschreibung
prestoql-connector-raptor	Ändern Sie die Werte in der Datei raptor.properties von Presto. SQL
prestoql-connector-redis	Ändern Sie die Werte in der Datei redis.properties von PrestoSQL.
prestoql-connector-redshift	Ändern Sie die Werte in der Datei redshift.properties von PrestoSQL.
prestoql-connector-tpch	Ändern Sie die Werte in der Datei tpch.properties von Presto. SQL
prestoql-connector-tpcds	Ändern Sie die Werte in der Datei tpcds.properties von PrestoSQL.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von Ranger. KMS
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 6.1.0

6.1.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for 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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

6.1.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.1.0. Änderungen beziehen sich auf Version 6.0.0.

Erste Version: 4. September 2020

Letzte Aktualisierung: 15. Oktober 2020

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.828
- Flink-Version 1.11.0
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBaseAusführung 2.2.5
- HBase-Operator-Tools 1.0.0
- HCatalogAusführung 3.1.2-amzn-0
- Hive-Version 3.1.2-amzn-1
- Hudi-Version 0.5.2-incubating
- Hue-Version 4.7.1
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0
- MXNetAusführung 1.6.0
- Oozie-Version 5.2.0

- Phoenix Version 5.0.0
- Presto-Version 0.232
- Presto, Version 3.38 SQL
- Spark-Version 3.0.0-amzn-0
- TensorFlow Ausführung 2.1.0
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

Neue Features

- ARM Instance-Typen werden ab EMR Amazon-Version 5.30.0 und EMR Amazon-Version 6.1.0 unterstützt.
- Allzweck-Instance-Typen von M6g werden ab den EMR Amazon-Versionen 6.1.0 und 5.30.0 unterstützt. Weitere Informationen finden Sie unter [Unterstützte Instance-Typen](#) im Amazon EMR Management Guide.
- Die EC2 Platzierungsgruppenfunktion wird ab EMR Amazon-Version 5.23.0 als Option für Cluster mit mehreren primären Knoten unterstützt. Derzeit werden nur Primärknoten-Typen von des Platzierungsgruppenfeatures unterstützt, und die SPREAD-Strategie wird auf diese Primärknoten angewendet. Bei dieser SPREAD-Strategie wird eine kleine Gruppe von Instances auf separater zugrundeliegender Hardware platziert, um den Verlust mehrerer Primärknoten im Falle eines Hardwarefehlers zu verhindern. Weitere Informationen finden Sie unter [EMR Integration mit EC2 Placement Group](#) im Amazon EMR Management Guide.
- Verwaltete Skalierung — Mit EMR Amazon-Version 6.1.0 können Sie Amazon EMR Managed Scaling aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster je nach Arbeitslast automatisch zu erhöhen oder zu verringern. Amazon wertet EMR kontinuierlich Cluster-Metriken aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster im Hinblick auf Kosten und Geschwindigkeit optimieren. Managed Scaling ist auch auf EMR Amazon-Version 5.30.0 und höher verfügbar, außer 6.0.0. Weitere Informationen finden Sie unter [Scaling Cluster Resources](#) im Amazon EMR Management Guide.
- Presto SQL Version 338 wird mit EMR 6.1.0 unterstützt. Weitere Informationen finden Sie unter [Presto](#).
 - Presto SQL wird nur in EMR 6.1.0 und späteren Versionen unterstützt, nicht in 6.0.0 oder 5.x.
EMR EMR

- Der Anwendungsname `Presto` wird weiterhin verwendet, um PrestoDB auf Clustern zu installieren. Verwenden Sie den Anwendungsnamen, um Presto SQL auf Clustern zu installieren. `PrestoSQL`
- Sie können entweder PrestoDB oder Presto installieren `SQL`, aber Sie können nicht beide auf einem einzigen Cluster installieren. Wenn beim Versuch, einen Cluster zu erstellen, sowohl PrestoDB als auch Presto angegeben `SQL` werden, tritt ein Validierungsfehler auf und die Anfrage zur Clustererstellung schlägt fehl.
- Presto `SQL` wird sowohl auf Single-Master- als auch auf Multi-Master-Clustern unterstützt. Auf Multi-Master-Clustern ist ein externer Hive-Metastore erforderlich, um Presto oder `SQL` PrestoDB auszuführen. Siehe [Unterstützte Anwendungen in einem EMR Cluster mit mehreren Primärknoten](#).
- ECR Unterstützung der auto Authentifizierung auf Apache Hadoop und Apache Spark mit Docker: Spark-Benutzer können Docker-Images von Docker Hub und Amazon Elastic Container Registry (Amazon ECR) verwenden, um Umgebungs- und Bibliotheksabhängigkeiten zu definieren.

[Konfigurieren Sie Docker](#) und [führen Sie Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x](#) aus.

- EMR unterstützt Apache ACID Hive-Transaktionen: Amazon EMR 6.1.0 bietet Unterstützung für ACID Hive-Transaktionen, sodass es den ACID Eigenschaften einer Datenbank entspricht. Mit diesem Feature können Sie `INSERT`, `UPDATE`, `DELETE`, - und `MERGE`-Operationen in Hive-Tabellen mit Daten in Amazon Simple Storage Service (Amazon S3) ausführen. Dies ist eine wichtige Funktion für Anwendungsfälle wie Streaming-Erfassung, Neuformulierung von Daten, Verwendung `MERGE` von Massenaktualisierungen und langsam ändernde Dimensionen. Weitere Informationen, einschließlich Konfigurationsbeispielen und Anwendungsfällen, finden Sie unter [Amazon EMR unterstützt Apache ACID Hive-Transaktionen](#).

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR

- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Apache Flink wird auf 6.0.0 nicht unterstützt, aber es wird auf 6.1.0 mit Flink 1.11.0 unterstützt. EMR Dies ist die erste Version von Flink, die Hadoop 3 offiziell unterstützt. Siehe Ankündigung der Veröffentlichung von [Apache Flink 1.11.0](#).
- Ganglia wurde aus den Standard-6.1.0-Paketpaketen entfernt. EMR

Bekannte Probleme

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren

Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine `ulimit`-Standard-Einstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffnete Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein `ulimit` über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein `ulimit` festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das `ulimit` für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
```

```
sudo mkdir -p /etc/systemd/system/$proc.service.d/  
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF  
[Service]  
LimitNOFILE=65536  
LimitNPROC=65536  
EOF  
pid=$(pgrep -f aws157.$proc.Main)  
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535  
done  
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich entscheidend auf alle Hudi-Einfüge-, Upsert- und Löschvorgänge auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi zu erhalten. RPM

- Wenn Sie eine benutzerdefinierte Garbage-Collection-Konfiguration mit `spark.driver.extraJavaOptions` und festlegen, führt dies dazu `spark.executor.extraJavaOptions`, dass der Start von Treiber/Executor mit 6.1 aufgrund einer widersprüchlichen Garbage-Collection-Konfiguration fehlschlägt. EMR In EMR Version 6.1.0 sollten Sie stattdessen eine benutzerdefinierte Spark-Garbage-Collection-Konfiguration für Treiber und Executors mit den Eigenschaften und angeben. `spark.driver.defaultJavaOptions` `spark.executor.defaultJavaOptions` Weitere Informationen finden Sie unter [Apache Spark Runtime Environment](#) und [Konfiguration von Spark Garbage Collection auf Amazon EMR 6.1.0](#).
- Die Verwendung von Pig mit Oozie (und innerhalb von Hue, da Hue Oozie-Aktionen zur Ausführung von Pig-Skripten verwendet) generiert den Fehler, dass eine native LZ0-Bibliothek nicht geladen werden kann. Diese Fehlermeldung ist informativ und verhindert nicht, dass Pig ausgeführt wird.
- Hudi-Parallelitätsunterstützung: Derzeit unterstützt Hudi keine gleichzeitigen Schreibvorgänge in eine einzelne Hudi-Tabelle. Darüber hinaus macht Hudi alle Änderungen rückgängig, die von in Bearbeitung befindlichen Autoren vorgenommen wurden, bevor ein neuer Writer beginnen kann. Gleichzeitige Schreibvorgänge können diesen Mechanismus stören und zu Wettlaufbedingungen führen, was zu Datenbeschädigungen führen kann. Sie sollten sicherstellen, dass im Rahmen Ihres Datenverarbeitungs-Workflows immer nur ein einziger Hudi-Writer mit einer Hudi-Tabelle arbeitet. Hudi unterstützt mehrere gleichzeitige Lesegeräte, die mit derselben Hudi-Tabelle arbeiten.

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- In Amazon EMR 6.1.0 gibt es ein Problem, das Cluster betrifft, auf denen Presto ausgeführt wird. Nach einem längeren Zeitraum (Tage) kann der Cluster Fehler wie „su: failed to execute /bin/bash: Resource temporarily unavailable“ oder „Shell-Anfrage auf Kanal 0 fehlgeschlagen“ ausgeben. Dieses Problem wird durch einen internen EMR Amazon-Prozess (InstanceController) verursacht, der zu viele Light-Weight-Prozesse (LWP) erzeugt, was letztendlich dazu führt, dass der Hadoop-Benutzer sein Nproc-Limit überschreitet. Dadurch wird verhindert, dass der Benutzer zusätzliche Prozesse öffnet. Die Lösung für dieses Problem ist ein Upgrade auf 6.2.0. EMR

6.1.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	2.0.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.42.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	3.2.1-amzn-1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httfs-server	3.2.1-amzn-1	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.2.1-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-1	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.2.5	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	2.2.5	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.5	HBase Befehlszeilenclient.
hbase-rest-server	2.2.5	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.5	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-2	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	3.1.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating-amzn-2	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi. SQL
hudi-spark	0.5.2-incubating-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0- -2.0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
prestoql-coordinator	338	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	338	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	338	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.0.0-amzn-0	In-Memory-Ausführungseengine fürYARN.

Komponente	Version	Beschreibung
spark-yarn-slave	3.0.0-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTP Apache-Server.
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

6.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-6.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> YARN von Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-env	Werte in der HDFS Umgebung ändern.
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. hive-exec-log
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
prestosql-node	Ändern Sie die Werte in der Datei SQL node.properties von Presto.
prestosql-connector-blackhole	Ändern Sie die Werte in der Datei blackhole.properties von PrestoSQL.
prestosql-connector-cassandra	Ändern Sie die Werte in der Datei cassandra.properties von PrestoSQL.
prestosql-connector-hive	Ändern Sie die Werte in der Datei hive.properties von Presto. SQL

Klassifizierungen	Beschreibung
prestosql-connector-jmx	Ändern Sie die Werte in der Datei <code>jmx.properties</code> von PrestoSQL.
prestosql-connector-kafka	Ändern Sie die Werte in der Datei <code>kafka.properties</code> von PrestoSQL.
prestosql-connector-localfile	Ändern Sie die Werte in der Datei <code>localfile.properties</code> von PrestoSQL.
prestosql-connector-memory	Ändern Sie die Werte in der Datei <code>memory.properties</code> von PrestoSQL.
prestosql-connector-mongodb	Ändern Sie die Werte in der Datei <code>mongodb.properties</code> von PrestoSQL.
prestosql-connector-mysql	Ändern Sie die Werte in der Datei <code>mysql.properties</code> von PrestoSQL.
prestosql-connector-postgresql	Ändern Sie die Werte in der Datei <code>postgresql.properties</code> von PrestoSQL.
prestosql-connector-raptor	Ändern Sie die Werte in der Datei <code>raptor.properties</code> von Presto. SQL
prestosql-connector-redis	Ändern Sie die Werte in der Datei <code>redis.properties</code> von PrestoSQL.
prestosql-connector-redshift	Ändern Sie die Werte in der Datei <code>redshift.properties</code> von PrestoSQL.
prestosql-connector-tpch	Ändern Sie die Werte in der Datei <code>tpch.properties</code> von Presto. SQL
prestosql-connector-tpcds	Ändern Sie die Werte in der Datei <code>tpcds.properties</code> von PrestoSQL.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei <code>dbks-site.xml</code> von Ranger. KMS

Klassifizierungen	Beschreibung
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 6.0.1

6.0.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for 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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

6.0.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR

- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPCE S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (\$region ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys](#).

6.0.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMRS3Select-Anschluss
emrfs	2.39.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-0.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-0.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-0.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	3.2.1-amzn-0.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-0.1	HTTPEndpunkt für OperationenHDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	3.2.1-amzn-0.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-0.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-0.1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	2.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.3	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	2.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	3.1.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-0	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.5.0-incubating-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0- -2.0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.230	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.230	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.230	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTPApache-Server.
zeppelin-server	0.9.0- SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

6.0.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-6.0.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> YARN von Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-env	Werte in der HDFS Umgebung ändern.
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat

Klassifizierungen	Beschreibung
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. hive-exec-log
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS

Klassifizierungen	Beschreibung
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Werte in der Umgebung ändern. YARN
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 6.0.0

6.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for 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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

6.0.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.0.0.

Erste Version: 10. März 2020

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.711
- Ganglia Version 3.7.2
- Hadoop Version 3.2.1

- HBaseAusführung 2.2.3
- HCatalogAusführung 3.1.2
- Hive Version 3.1.2
- Hudi Version 0.5.0-incubating
- Hue Version 4.4.0
- JupyterHub Ausführung 1.0.0
- Livy Version 0.6.0
- MXNetAusführung 1.5.1
- Oozie Version 5.1.0
- Phoenix Version 5.0.0
- Presto Version 0.230
- Spark Version 2.4.4
- TensorFlow Ausführung 1.14.0
- Zeppelin versie 0.9.0- SNAPSHOT
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

Note

Flink, Sqoop, Pig und Mahout sind in der EMR Amazon-Version 6.0.0 nicht verfügbar.

Neue Features

- YARN Docker Runtime Support — YARN Anwendungen wie Spark-Jobs können jetzt im Kontext eines Docker-Containers ausgeführt werden. Auf diese Weise können Sie auf einfache Weise Abhängigkeiten in einem Docker-Image definieren, ohne benutzerdefinierte Bibliotheken auf Ihrem EMR Amazon-Cluster installieren zu müssen. Weitere Informationen finden [Sie unter Docker-Integration konfigurieren](#) und [Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.0.0 ausführen](#).
- LLAPHive-Unterstützung — Hive unterstützt jetzt den LLAP Ausführungsmodus für eine verbesserte Abfrageleistung. Weitere Informationen finden Sie unter Hive [verwenden](#). LLAP

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Amazon Linux
 - Amazon Linux 2 ist das Betriebssystem für die EMR 6.x-Release-Serie.
 - `systemd` wird anstelle von `in` Amazon Linux 1 für das Servicemanagement `upstart` verwendet.
- Java-Entwicklungskit (JDK)

- Corretto JDK 8 ist der Standard JDK für die EMR 6.x-Release-Serie.
- Scala
 - Scala 2.12 wird mit Apache Spark und Apache Livy verwendet.
- Python 3
 - Python 3 ist jetzt die Standardversion von Python in EMR.
- YARN Knotenbeschriftungen
 - Ab der Amazon EMR 6.x-Release-Serie ist die Funktion YARN Node Labels standardmäßig deaktiviert. Die Anwendungs-Master-Prozesse können standardmäßig sowohl auf Kern- als auch auf Aufgabenknoten ausgeführt werden. Sie können die Funktion für YARN Knotenbeschriftungen aktivieren, indem Sie die folgenden Eigenschaften konfigurieren: `yarn.node-labels.enabled` und `yarn.node-labels.am.default-node-label-expression`. Weitere Informationen finden Sie unter [Grundlegendes zu Primär-, Kern- und Aufgabenknoten](#).

Bekanntes Problem

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- Die interaktive Spark-Shell PySpark, einschließlich SparkR und Spark-Shell, unterstützt die Verwendung von Docker mit zusätzlichen Bibliotheken nicht.

- Um Python 3 mit Amazon EMR Version 6.0.0 zu verwenden, müssen Sie Folgendes PATH hinzufügen `yarn.nodemanager.env-whitelist`.
- Die Funktionen Live Long und Process (LLAP) werden nicht unterstützt, wenn Sie den AWS Glue-Datenkatalog als Metastore für Hive verwenden.
- Wenn Sie Amazon EMR 6.0.0 mit Spark- und Docker-Integration verwenden, müssen Sie die Instances in Ihrem Cluster mit demselben Instance-Typ und derselben Anzahl von EBS Volumes konfigurieren, um Fehler beim Senden eines Spark-Jobs mit Docker-Laufzeit zu vermeiden.
- In Amazon EMR 6.0.0 ist der Speichermodus HBase auf Amazon S3 vom Problem [HBASE-24286](#) betroffen. HBaseMaster kann nicht initialisiert werden, wenn der Cluster mit vorhandenen S3-Daten erstellt wird.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

6.0.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMRS3Select-Anschluss
emrfs	2.39.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-0	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-0	HDFSBefehlszeilenclient und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.2.1-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-0	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-0	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	3.2.1-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	3.2.1-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-0	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-0	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	2.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	2.2.3	HBase Befehlszeilenclient.
hbase-rest-server	2.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	2.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	3.1.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	3.1.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	3.1.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.0-incubating-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0- -2.0 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0- -2,0 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.230	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.230	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.230	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.41+	HTTPApache-Server.
zeppelin-server	0.9.0- SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

6.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-6.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> YARN von Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-env	Werte in der HDFS Umgebung ändern.
hdfs-site	Ändern Sie die Werte in HDFS der Datei hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.

Klassifizierungen	Beschreibung
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive	Amazon EMR — kuratierte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. hive-exec-log
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie die Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der Umgebung ändern. YARN
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

Amazon EMR 5.x-Release-Versionen

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon EMR 5.x-Release-Version verfügbar sind.

Wenn Sie einen Cluster starten, können Sie aus mehreren Versionen von Amazon wählenEMR. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den

Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.2.0`.

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon EMR 5.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#).

Themen

- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [EMRAmazon-Version 5.36.2](#)
- [EMRAmazon-Version 5.36.1](#)
- [EMRAmazon-Version 5.36.0](#)
- [EMRAmazon-Version 5.35.0](#)
- [EMRAmazon-Version 5.34.0](#)
- [EMRAmazon-Version 5.33.1](#)
- [EMRAmazon-Version 5.33.0](#)
- [EMRAmazon-Version 5.32.1](#)
- [EMRAmazon-Version 5.32.0](#)
- [EMRAmazon-Version 5.31.1](#)
- [EMRAmazon-Version 5.31.0](#)
- [EMRAmazon-Version 5.30.2](#)
- [EMRAmazon-Version 5.30.1](#)
- [EMRAmazon-Version 5.30.0](#)
- [EMRAmazon-Version 5.29.0](#)
- [EMRAmazon-Version 5.28.1](#)
- [EMRAmazon-Version 5.28.0](#)
- [EMRAmazon-Version 5.27.1](#)
- [EMRAmazon-Version 5.27.0](#)

- [EMRAmazon-Version 5.26.0](#)
- [EMRAmazon-Version 5.25.0](#)
- [EMRAmazon-Version 5.24.1](#)
- [EMRAmazon-Version 5.24.0](#)
- [EMRAmazon-Version 5.23.1](#)
- [EMRAmazon-Version 5.23.0](#)
- [EMRAmazon-Version 5.22.0](#)
- [EMRAmazon-Version 5.21.2](#)
- [EMRAmazon-Version 5.21.1](#)
- [EMRAmazon-Version 5.21.0](#)
- [EMRAmazon-Version 5.20.1](#)
- [EMRAmazon-Version 5.20.0](#)
- [EMRAmazon-Version 5.19.1](#)
- [EMRAmazon-Version 5.19.0](#)
- [EMRAmazon-Version 5.18.1](#)
- [EMRAmazon-Version 5.18.0](#)
- [EMRAmazon-Version 5.17.2](#)
- [EMRAmazon-Version 5.17.1](#)
- [EMRAmazon-Version 5.17.0](#)
- [EMRAmazon-Version 5.16.1](#)
- [EMRAmazon-Version 5.16.0](#)
- [EMRAmazon-Version 5.15.1](#)
- [EMRAmazon-Version 5.15.0](#)
- [EMRAmazon-Version 5.14.2](#)
- [EMRAmazon-Version 5.14.1](#)
- [EMRAmazon-Version 5.14.0](#)
- [EMRAmazon-Version 5.13.1](#)
- [EMRAmazon-Version 5.13.0](#)
- [EMRAmazon-Version 5.12.3](#)

- [EMRAmazon-Version 5.12.2](#)
- [EMRAmazon-Version 5.12.1](#)
- [EMRAmazon-Version 5.12.0](#)
- [EMRAmazon-Version 5.11.4](#)
- [EMRAmazon-Version 5.11.3](#)
- [EMRAmazon-Version 5.11.2](#)
- [EMRAmazon-Version 5.11.1](#)
- [EMRAmazon-Version 5.11.0](#)
- [EMRAmazon-Version 5.10.1](#)
- [EMRAmazon-Version 5.10.0](#)
- [EMRAmazon-Version 5.9.1](#)
- [EMRAmazon-Version 5.9.0](#)
- [EMRAmazon-Version 5.8.3](#)
- [EMRAmazon-Version 5.8.2](#)
- [EMRAmazon-Version 5.8.1](#)
- [EMRAmazon-Version 5.8.0](#)
- [EMRAmazon-Version 5.7.1](#)
- [EMRAmazon-Version 5.7.0](#)
- [EMRAmazon-Version 5.6.1](#)
- [EMRAmazon-Version 5.6.0](#)
- [EMRAmazon-Version 5.5.4](#)
- [EMRAmazon-Version 5.5.3](#)
- [EMRAmazon-Version 5.5.2](#)
- [EMRAmazon-Version 5.5.1](#)
- [EMRAmazon-Version 5.5.0](#)
- [EMRAmazon-Version 5.4.1](#)
- [EMRAmazon-Version 5.4.0](#)
- [EMRAmazon-Version 5.3.2](#)
- [EMRAmazon-Version 5.3.1](#)

- [EMRAmazon-Version 5.3.0](#)
- [EMRAmazon-Version 5.2.3](#)
- [EMRAmazon-Version 5.2.2](#)
- [EMRAmazon-Version 5.2.1](#)
- [EMRAmazon-Version 5.2.0](#)
- [EMRAmazon-Version 5.1.1](#)
- [EMRAmazon-Version 5.1.0](#)
- [EMRAmazon-Version 5.0.3](#)
- [EMRAmazon-Version 5.0.2](#)
- [EMRAmazon-Version 5.0.1](#)
- [EMRAmazon-Version 5.0.0](#)

Anwendungsversionen in Amazon-EMR-5.x-Versionen (PNG)

Eine umfassende Tabelle mit den in jeder Amazon-EMR-5.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-5.x-Versionen](#) in Ihrem Browser.

EMRAmazon-Version 5.36.2

5.36.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
AWS SDK for 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
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

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
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
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

Versionshinweise zu 5.36.2

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.36.2. Die Änderungen beziehen sich auf 5.36.1. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Amazon die Kernknoten EMR nicht unter die für den Cluster eingestellte HDFS Replikationsfaktor-Einstellung herunterskaliert.

Diese Verbesserung erfüllt die Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.

- Diese Version fügt dem Cluster-Skalierungs-Workflow für die Ausführung Presto oder Trino einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung reduziert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit ausgeführt wird. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Behebt ein Problem, bei dem Cluster-Scale-Down-Operationen möglicherweise zum Stillstand kommen, während Amazon EMR einen Core-Node ordnungsgemäß außer Betrieb nimmt und dieser vor der vollständigen Außerbetriebnahme fehlerhaft wird.
- Verbessert die Stabilität eines Knotens in einem Hochverfügbarkeitscluster mit mehreren Primärknoten, wenn Amazon einen einzelnen Knoten EMR neu startet.
- Optimiert die Protokollverwaltung, wenn Amazon auf Amazon EMR EC2 läuft. Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und den Clusterbetrieb unterbrechen.
- Behebt einen seltenen Fehler, der dazu führen kann, dass ein Hochverfügbarkeitscluster mit mehreren Primärknoten ausfällt, weil er nicht mit dem Yarn kommunizieren kann.
ResourceManager
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als Standard-Amazon EMRAMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 503.0	4,14.343	xxxxxx, 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

5.36.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.16.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.21.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMRS3Select-Anschluss
emrfs	2.51.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-4	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	2.10.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.10.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.9-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.9-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.9-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.10.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.10.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.
nvidia-cuda	11.0.194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.267-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.8-amzn-2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.8-amzn-2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.36.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.36.2-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Hadoop-Datei <code>container-executor.cfg</code> . YARN	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei <code>server.properties</code> .	Restarts EMR record server.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer <code>log4j.properties</code> .	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der <code>spark-defaults.conf</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der <code>metrics.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei <code>oraoop-site.xml</code> OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

5.36.2 Änderungsprotokoll

Änderungsprotokoll für Version 5.36.2 und Versionshinweise

Datum	Ereignis	Beschreibung
2024-05-28	Die Bereitstellung ist abgeschlossen	Amazon EMR 5.36.2 vollständig in allen unterstützten Regionen bereitgestellt
2024-05-28	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 5.36.2
20.05.2024-05	Erstversion	Amazon EMR 5.36.2 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

EMRAmazon-Version 5.36.1

5.36.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
AWS SDK for 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
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

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,267	0,266	0,261
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.36.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.36.1. Änderungen beziehen sich auf Version 5.36.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Die EMR Amazon-Version 5.36.1 bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Scale-Down. In früheren 5.x-Versionen konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die Verbesserung stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des

Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).

- Die Version 5.36.1 verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem Cluster. EMR Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Die Version 5.36.1 startet den Log-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Die Version 5.36.1 behebt ein Problem, bei dem EMR Amazon-Daemons auf dem primären Knoten veraltete Metadaten für beendete Instances im Cluster beibehielten. Die Beibehaltung veralteter Daten kann dazu führen, dass die Nutzung innerhalb des Clusters CPU und des Speichers unbegrenzt zunimmt, was letztendlich zu Clusterausfällen führen kann.
- Bei Clustern, die mit mehreren Primärknoten gestartet werden, behebt die Version 5.36.1 ein Problem, bei dem ein EC2 Amazon-Hardwarefehler auf einem der primären Knoten zum Ausfall eines zweiten Primärknotens führen und Ihren Cluster instabil machen könnte.
- Für Cluster, die mit Verschlüsselung während der Übertragung konfiguriert sind, unterstützt verwaltete Skalierung jetzt die Datenmenge von Spark Shuffle. Spark-Shuffle-Daten sind Daten, die Spark auf mehrere Partitionen verteilt, um bestimmte Operationen auszuführen. Beim Herunterskalieren ignoriert verwaltete Skalierung die Instances mit Shuffle-Daten. Dadurch werden wiederholte Auftragsversuche und Neuberechnungen verhindert, was sich negativ auf den Preis und die Leistung auswirkt. [Weitere Informationen zu Shuffle-Vorgängen finden Sie im Spark-Programmierhandbuch](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als Standard-AZ Amazon EMRAMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906,0		04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

5.36.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.16.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.21.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMRS3Select-Anschluss
emrfs	2.51.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httfs-server	2.10.1-amzn-4	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.10.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.9-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.9-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	2.3.9-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.10.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.10.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	11.0.194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.267-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-2	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.4.8-amzn-2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Clients.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.36.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.36.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die <code>flink-conf.yaml</code> -Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink <code>log4j-session.properties</code> für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei <code>server.properties</code> .	Restarts EMR record server.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer <code>log4j.properties</code> .	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der <code>spark-defaults.conf</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der <code>metrics.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei <code>oraoop-site.xml</code> OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

5.36.1 Änderungsprotokoll

Änderungsprotokoll für Version 5.36.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
2023-05-25	Die Bereitstellung ist abgeschlossen	Amazon EMR 5.36.1 vollständig in allen unterstützten Regionen bereitgestellt
2023-05-09	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 5.36.1
2023-05-04	Erstversion	Amazon EMR 5.36.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

EMRAmazon-Version 5.36.0

5.36.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
AWS SDK for 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
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	-	-	-	-

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
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
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

5.36.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.36.0. Änderungen beziehen sich auf Version 5.35.0.

Erste Version: 15. Juni 2022

Neue Features

- EMRAmazon-Version 5.36.0 bietet Unterstützung für Data Definition Language (DDL) mit Apache Spark auf Apache Ranger-fähigen Clustern. Auf diese Weise können Sie Apache Ranger für die Verwaltung des Zugriffs für Operationen wie das Erstellen, Ändern und Löschen von Datenbanken und Tabellen aus einem EMR Amazon-Cluster verwenden.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als Standard-Amazon EMRAMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API oder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Asien-Pazifik (Melbourne), Israel (Tel Aviv)
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004.0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 719.0	4,14.287	10. August 2022	USA West (Nordkalifornien), Europa (Paris), Europa (Stockholm), Europa (Frankfurt), Asien-Pazifik (Mumbai), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 426,0	4,14.281	14. Juni 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR 5.36.0-Upgrades unterstützen jetzt: 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, Amazon Glue-Konnektor 1.18.0, 2.51.0. EMRFS

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei RecordServer Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. SecretAgent Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

5.36.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.16.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.21.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMRS3Select-Anschluss
emrfs	2.51.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.10.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-4	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	2.10.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.10.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.10.1-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-2	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.9-amzn-2	Bereitstellung von DienstenH-Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.9-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.9-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.10.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.10.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.
nvidia-cuda	11.0.194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.267-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-2	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.4.8-amzn-2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.36.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.36.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 5.35.0

5.35.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
AWS SDK for 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
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

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
Hudi	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,266	0,261	0,245,1	0,245,1
Spark	2.4.8	2.4.8	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
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

5.35.0 Versionshinweise

Dies ist der EMR Versionshinweis zu Amazon-Version 5.35.0.

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.35.0. Änderungen beziehen sich auf Version 5.34.0.

Erste Version: 30. März 2022

Neue Features

- Amazon-Anwendungen der EMR Version 5.35, die Log4j 1.x und Log4j 2.x verwenden, werden auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert und erfordern keine Bootstrap-Aktionen zur Behebung der Probleme in früheren Versionen. CVE Siehe [Ansatz zur Minderung CVE -2021-44228](#).

Änderungen, Verbesserungen und behobene Probleme

Änderungen bei Flink

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> • Aktualisieren Sie die Flink-Version auf 1.14.2. • log4j wurde auf 2.17.1 aktualisiert.

Hadoop-Änderungen

Änderungstyp	Beschreibung
Open-Source-Backports für Hadoop seit 5.34.0 EMR	<ul style="list-style-type: none"> • YARN-10438: Handle Null in C # () containerId lientRMService getContainerReport • YARN-7266: Threads für den Timeline-Server-Event-Handler sind gesperrt • YARN-10438: ATS 1.5 kann nicht gestartet werden, wenn RollingLevelDb Dateien beschädigt sind oder fehlen

Änderungstyp	Beschreibung
	<ul style="list-style-type: none"> • HADOOP-13500: Synchronisieren der Iteration des Objekts mit den Konfigurationseigenschaften • YARN-10651: Ist mit in abgestürzt. CapacityScheduler NPE AbstractYarnScheduler updateNodeResource() • HDFS-12221: Ersetzen Sie Xerces in XmlEditsVisitor • HDFS-16410: Unsicheres Xml-Parsen bei OfflineEditsXMLLoader
Hadoop-Änderungen und Korrekturen	<ul style="list-style-type: none"> • Tomcat wird in verwendet und HttpFS wurde auf KMS 8.5.75 aktualisiert • In FileSystemOptimizedCommitter V2 wurde die Erfolgsmarkierung in den commitJob Ausgabepfad geschrieben, der bei der Erstellung des Committers definiert wurde. Da Ausgabepfade commitJob und Ausgabepfade auf Aufgabenebene unterschiedlich sein können, wurde der Pfad korrigiert, sodass er den in den Manifestdateien definierten Pfad verwendet. Bei Hive-Jobs führt dies dazu, dass die Erfolgsmarkierung korrekt geschrieben wird, wenn Operationen wie dynamische Partition oder UNION ALL ausgeführt werden.

Änderungen bei Hive

Änderungstyp	Beschreibung
Hive wurde auf die Open-Source-Version 2.3.9 aktualisiert, einschließlich dieser Korrekturen JIRA	<ul style="list-style-type: none"> • HIVE-17155: findConfFile () in HiveConf .java hat einige Probleme mit dem Conf-Pfad

Änderungstyp	Beschreibung
	<ul style="list-style-type: none"> • HIVE-24797: Deaktiviert die Validierung von Standardwerten beim Parsen von Avro-Schemas • HIVE-21563: Verbessern Sie die Leistung von Table#, indem Sie Once deaktivieren getEmptyTable registerAllFunctions • HIVE-18147: Tests können mit java.net fehlschlagen. BindException: Adresse wird bereits verwendet • HIVE-24608: Wechselt im Client für Hive 2.3.x zurück zu get_table HMS • HIVE-21200: Vektorisierung — Die Datumsspalte wirft java.lang aus. UnsupportedOperationException für Parkett • HIVE-19228: Entfernen Sie die Verwendung von commons-httpclient 3.x
Hive Open-Source-Backports seit 5.34.0 EMR	<ul style="list-style-type: none"> • HIVE-19990: Eine Abfrage mit einem Intervall literal in der Join-Bedingung schlägt fehl • HIVE-25824: Aktualisieren Sie Branch-2.3 auf log4j 2.17.0 • TEZ-4062: Die Planung spekulativer Versuche sollte abgebrochen werden, wenn die Aufgabe abgeschlossen ist • TEZ-4108: Während eines spekulativen Ausführungsrennens NullPointerException • TEZ-3918: Die Einstellung tez.task.log.level funktioniert nicht

Änderungstyp	Beschreibung
Hive-Upgrades und -Korrekturen	<ul style="list-style-type: none"> • Aktualisieren Sie die Log4j-Version auf 2.17.1 • Aktualisieren ORC Sie die Version auf 1.4.3 • Der Deadlock aufgrund eines fehlerhaften Thread-Ins wurde behoben ShuffleScheduler
Neue Features	<ul style="list-style-type: none"> • Feature zum Drucken von Hive-Abfragen in AM-Protokollen hinzugefügt. Diese ist standardmäßig deaktiviert. Flag/Konf.: <code>tez.am.emr.print.hive.query.in.log</code> Status (Standard):FALSE.

Oozie-Änderungen

Änderungstyp	Beschreibung
Oozie bietet seit 5.34.0 Backports auf Open-Source-Basis EMR	<ul style="list-style-type: none"> • OOZIE-3652: Der Oozie-Launcher sollte die Verzeichnisliste erneut versuchen, wenn dies der Fall ist NoSuchFileException

Pig-Änderungen

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> • log4j wurde auf 1.2.17 aktualisiert.

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.

- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.35.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.15.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.20.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMRS3Select-Anschluss
emrfs	2.49.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR Knoten für Apache Flink. JobManager
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-3	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-3	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-3	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.10.1-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.10.1-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.10.1-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.9-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.9-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.9-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu HCatalog bereitstellt.
hive-client	2.3.9-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.9-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.9.0-amzn-2	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.9.0-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.

Komponente	Version	Beschreibung
hudi-presto	0.9.0-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.266-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.266-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.266-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.8-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.8-amzn-1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.35.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.35.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei log4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei <code>server.properties</code> .	Restarts EMR record server.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer <code>log4j.properties</code> .	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der <code>spark-defaults.conf</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der <code>metrics.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei <code>oraoop-site.xml</code> OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 5.34.0

5.34.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
AWS SDK for 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

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
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
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
Jupyter Enterprise Gateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,261	0,245,1	0,245,1	0,240,1
Spark	2.4.8	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.9.0	0.9.0	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.34.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.34.0. Änderungen beziehen sich auf Version 5.33.1.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

Neue Features

- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).
- [Hudi] Verbesserungen zur Vereinfachung der Hudi-Konfiguration. Optimistische Parallelitätssteuerung standardmäßig deaktiviert.

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Bisher führte ein manueller Neustart des Resource Managers auf einem Multi-Master-Cluster dazu, dass EMR Amazon-On-Cluster-Daemons wie Zookeeper alle zuvor außer Betrieb genommenen oder verloren gegangenen Knoten in der Zookeeper-Znode-Datei neu luden. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon entfernt

EMR jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei und die internen Grenzwerte wurden erhöht.

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Zeppelin auf Version 0.10.0 upgegradet.
- Livy Fix – auf 0.7.1 aktualisiert
- Leistungsverbesserung bei Spark — Heterogene Executors werden deaktiviert, wenn bestimmte Spark-Konfigurationswerte in 5.34.0 außer Kraft gesetzt werden. EMR
- Web HDFS - und HttpFS-Server sind standardmäßig deaktiviert. Sie können Web HDFS mithilfe der Hadoop-Konfiguration erneut aktivieren, `dfs.webhdfs.enabled` Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

Bekannte Probleme

- Die Amazon EMR Notebooks-Funktion, die mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den der Livy-Identitätswechsel aktiviert ist. Die Problemumgehung besteht darin, den HttpFS-Server zu starten, bevor Sie das Notebook mit dem EMR Cluster verbinden. `sudo systemctl start hadoop-https`
- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache Hadoop HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem EMR primären Amazon-Node entweder manuell mithilfe `sudo systemctl start hadoop-https` oder [verwenden Sie einen](#) Amazon-Schritt. EMR
- Die Amazon EMR Notebooks-Funktion, die mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den der Livy-Identitätswechsel aktiviert ist. Die Problemumgehung besteht darin, den HttpFS-Server zu starten, bevor Sie das Notebook mit dem EMR Cluster verbinden. `sudo systemctl start hadoop-https`
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.34.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.14.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.4.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
<code>emr-s3-dist-cp</code>	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	1.7.0	EMRS3Select-Anschluss
emrfs	2.48.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.13.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.13.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-2	HDFSDienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.10.1-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-2	HDFS Dienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-https-server	2.10.1-amzn-2	HTTP Endpunkt für Operationen in HDFS.
hadoop-kms-server	2.10.1-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.10.1-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.8-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.8-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.8-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.8-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.8-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.8-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.8-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.9.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.9.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.9.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.261-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.261-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.261-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.8-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.34.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.34.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	Restarts the Hadoop HDFS services Namenode, Secondary Namenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-orahome-site	Ändern Sie die Werte in der Datei orahome-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 5.33.1

5.33.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
AWS SDK for 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

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
Hudi	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,245,1	0,240,1	0,240,1
Spark	2.4.7	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
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

5.33.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.33.0/5.33.1. Änderungen beziehen sich auf Version 5.32.0.

Erste Version: 19. April 2021

Letzte Aktualisierung: 9. August 2021

Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.15.0 aktualisiert
- AWS SDK for Java Auf Version 1.11.970 aktualisiert
- Auf Version 2.46.0 aktualisiert EMRFS
- EMRGoodies auf Version 2.14.0 aktualisiert
- EMRRecord Server wurde auf Version 1.9.0 aktualisiert
- EMRS3 Dist CP wurde auf Version 2.18.0 aktualisiert
- EMRSecret Agent wurde auf Version 1.8.0 aktualisiert
- Flink auf Version 1.12.1 aktualisiert
- Hadoop wurde auf Version 2.10.1-amzn-1 aktualisiert
- Upgrade von Hive auf Version 2.3.7-amzn-4
- Hudi wurde auf Version 0.7.0 aktualisiert
- Hue auf Version 4.9.0 aktualisiert
- OpenCV wurde auf Version 4.5.0 aktualisiert
- Presto auf Version 0.245.1-amzn-0 aktualisiert
- R wurde auf Version 4.0.2 aktualisiert
- Spark auf Version 2.4.7-amzn-1 aktualisiert
- Auf Version TensorFlow 2.4.1 aktualisiert
- Upgrade von Zeppelin auf Version 0.9.0

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Konfiguration eines Clusters zur Behebung von Leistungsproblemen mit Apache Timeline Server Version 1 und 1.5 YARN

Apache YARN Timeline Server Version 1 und 1.5 können bei sehr aktiven, großen EMR Clustern zu Leistungsproblemen führen, insbesondere bei `beiyarn.resourcemanager.system-metrics-publisher.enabled=true`, was die Standardeinstellung in Amazon ist EMR. Ein YARN Open-Source-Timeline Server v2 löst das Leistungsproblem im Zusammenhang mit der Skalierbarkeit von YARN Timeline Server.

Weitere Lösungen für dieses Problem umfassen:

- Konfiguration von `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- Aktivieren des Fixes für dieses Problem beim Erstellen eines Clusters, wie unten beschrieben.

Die folgenden EMR Amazon-Versionen enthalten eine Lösung für dieses Leistungsproblem mit dem YARN 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

Um den Fix für eine der oben angegebenen EMR Amazon-Versionen zu aktivieren, legen Sie diese Eigenschaften `true` in einer JSON Konfigurationsdatei fest, die mit dem [aws emr create-cluster](#) Befehlsparameter übergeben wird: `--configurations file://./configurations.json`. Oder aktivieren Sie den Fix über die Benutzeroberfläche der [Rekonfigurationskonsole](#).

Beispiel für den Inhalt der Datei `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": []
}
```

- Die Spark-Laufzeit ist jetzt schneller, wenn Partitionsspeicherorte aus Hive Metastore für Spark-Insert-Abfragen abgerufen werden.
- Aktualisierte Komponentenversionen. Eine Liste der Komponentenversionen finden Sie unter [About Amazon EMR Releases](#) in diesem Handbuch.
- Das AWS SDK Java-Bundle wurde auf jedem neuen Cluster installiert. Dies ist ein einzelnes JAR, das alle Dienste SDKs und ihre Abhängigkeiten enthält, anstatt einzelne Komponenten-Jars. Weitere Informationen finden Sie unter [Java SDK Bundled Dependency](#).
- Managed Scaling-Probleme in früheren EMR Amazon-Versionen wurden behoben und Verbesserungen vorgenommen, sodass die Ausfallraten von Anwendungen erheblich reduziert wurden.

- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositoryn aktiviert. Wenn Sie eine Amazon S3 VPC S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositoryn.](#)

Neue Features

- Amazon EMR unterstützt Amazon S3 Access Points, eine Funktion von Amazon S3, mit der Sie den Zugriff auf gemeinsam genutzte Data Lakes einfach verwalten können. Mit Ihrem Amazon S3 Access Point-Alias können Sie Ihren Datenzugriff auf Amazon in großem Umfang vereinfachen. Sie können Amazon S3 Access Points mit allen Versionen von Amazon EMR ohne zusätzliche Kosten in allen AWS Regionen verwenden, in denen Amazon verfügbar ist. Weitere Informationen zu Amazon S3 Access Points und Zugangspunkt-Aliassen finden Sie unter [Verwenden eines Alias im Bucket-Stil für Ihren Zugangspunkt](#) im Amazon-S3-Benutzerhandbuch.
- Amazon EMR 5.33 unterstützt neue EC2 Amazon-Instanz-Typen: c5a, c5ad, c6gn, c6gd, m6gd, d3, d3en, m5zn, r5b, r6gd. Siehe [Unterstützte Instanz-Typen](#).

Bekannte Probleme

- Niedrigeres Limit für AL2 die maximale Anzahl geöffneter Dateien bei älteren Versionen [in neueren Versionen behoben]. EMR Amazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standardeinstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt

Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Websites wie SparkUIs, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.

Important

EMRCluster, auf denen Amazon Linux oder Amazon Linux 2 Amazon Machine Images (AMIs) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen EC2 Amazon-Instances, auf denen das standardmäßige Amazon Linux ausgeführt wird. Wenn neue Amazon Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel und CUDA Updates) NVIDIA, verfügbar werden, nachdem eine EMR Amazon-Version verfügbar wird, laden EMR Cluster-Instances, die standardmäßig ausgeführt werden, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon so anpassen EMR AMI](#), dass es [das neueste Amazon Linux verwendet AMI](#).

- Die Konsolenunterstützung zur Erstellung einer Sicherheitskonfiguration, die die AWS Ranger-Integrationsoption spezifiziert, wird derzeit in der GovCloud Region nicht unterstützt. Die Sicherheitskonfiguration kann mit dem CLI vorgenommen werden. Weitere Informationen finden Sie unter [EMR Sicherheitskonfiguration erstellen](#) im Amazon EMR Management Guide.
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, EMR hat Amazon EMR verwaltete Standardrichtlinien mit Geltungsbereich v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [EMR Von Amazon verwaltete Richtlinien](#).
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.

- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.33.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.14.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMRS3Select-Anschluss
emrfs	2.46.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-1.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-1.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-1.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-1.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-1.1	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.10.1-amzn-1.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.10.1-amzn-1.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-1.1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.7-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-4	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstellt H Catalog.
hive-client	2.3.7-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-4	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.7-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.7.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.7.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.

Komponente	Version	Beschreibung
hudi-presto	0.7.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	Mein DatenbankserverSQL.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-1.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.7-amzn-1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.7-amzn-1.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.7-amzn-1.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.33.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.33.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei log4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei <code>server.properties</code> .	Restarts EMR record server.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer <code>log4j.properties</code> .	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der <code>spark-defaults.conf</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der <code>metrics.properties</code> -Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei <code>oraoop-site.xml</code> OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 5.33.0

5.33.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
AWS SDK for 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
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

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
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
Presto	0,245,1	0,240,1	0,240,1	0,238,3
Spark	2.4.7	2.4.7	2.4.7	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.33.0 Versionshinweise

5.33.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.14.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMRS3Select-Anschluss
emrfs	2.46.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-1	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.10.1-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.10.1-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.10.1-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBaseBefehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-4	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.7-amzn-4	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.7-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-4	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.7-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.7.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.7.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.7.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.7-amzn-1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.7-amzn-1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.33.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.33.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei <code>container-executor.cfg</code> von Hadoop YARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei <code>container-log4j.properties</code> von Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in 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

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 5.32.1

5.32.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
AWS SDK for 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

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
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

5.32.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPCE S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-`

`$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys.](#)

Bekanntete Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.32.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-

Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.1.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.17.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMRS3Select-Anschluss
emrfs	2.45.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-0.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-0.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-0.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-0.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.10.1-amzn-0.1	HTTP-Endpoint für Operationen in HDFS.
hadoop-kms-server	2.10.1-amzn-0.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.10.1-amzn-0.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.10.1-amzn-0.1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-3	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-3	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.7-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-3	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-3	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.7-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	Mein DatenbankserverSQL.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit

Komponente	Version	Beschreibung
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.240.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.240.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.240.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.7-amzn-0.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.7-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.32.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.32.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFSEinstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMRAmazon-Version 5.32.0

5.32.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
AWS SDK for 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

5.32.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.32.0. Änderungen beziehen sich auf Version 5.31.0.

Erste Version: 8. Januar 2021

Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.14.0 aktualisiert
- Amazon SageMaker Spark wurde SDK auf Version 1.4.1 aktualisiert
- Auf Version AWS SDK for Java 1.11.890 aktualisiert
- EMRDynamoDB Connector Version 4.16.0 wurde aktualisiert
- Auf Version 2.45.0 aktualisiert EMRFS
- EMRLog Analytics-Metriken wurden auf Version 1.18.0 aktualisiert
- EMR MetricsAndEventsApiGateway Der Client wurde auf Version 1.5.0 aktualisiert
- EMRRecord Server wurde auf Version 1.8.0 aktualisiert
- EMRS3 Dist CP wurde auf Version 2.17.0 aktualisiert
- EMRSecret Agent wurde auf Version 1.7.0 aktualisiert
- Flink auf Version 1.11.2 aktualisiert
- Hadoop wurde auf Version 2.10.1-amzn-0 aktualisiert
- Upgrade von Hive auf Version 2.3.7-amzn-3
- Hue auf Version 4.8.0 aktualisiert
- Mxnet wurde auf Version 1.7.0 aktualisiert
- OpenCV wurde auf Version 4.4.0 aktualisiert
- Presto auf Version 0.240.1-amzn-0 aktualisiert
- Spark auf Version 2.4.7-amzn-0 aktualisiert
- Auf Version TensorFlow 2.3.1 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Aktualisierte Komponentenversionen.
- Eine Liste der Komponentenversionen finden Sie unter [About Amazon EMR Releases](#) in diesem Handbuch.

Neue Features

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Sizierung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.

- **Unterstützungsstatus von Instance Metadata Service (IMDS) V2:** Die Komponenten Amazon EMR 5.23.1, 5.27.1 und 5.32 oder höher werden für alle Aufrufe verwendet. IMDSv2 IMDS Für IMDS Aufrufe in Ihrem Anwendungscode können Sie sowohl als auch IMDSv1 verwenden oder die Option so konfigurieren IMDSv2, dass sie nur für zusätzliche Sicherheit IMDS verwendet wird. IMDSv2 Bei anderen EMR 5.x-Versionen IMDSv1 führt die Deaktivierung zu einem Fehler beim Starten des Clusters.
- **Ab Amazon EMR 5.32.0** können Sie einen Cluster starten, der nativ in Apache Ranger integriert ist. Apache Ranger ist ein Open-Source-Framework zur Aktivierung, Überwachung und Verwaltung einer umfassenden Datensicherheit auf der gesamten Hadoop-Plattform. Weitere Informationen finden Sie unter [Apache Ranger](#). Dank der nativen Integration können Sie Ihren eigenen Apache Ranger verwenden, um eine detaillierte Datenzugriffskontrolle auf Amazon durchzusetzen. EMR Weitere Informationen finden Sie unter [Integrieren von Amazon EMR mit Apache Ranger](#) im EMR Amazon-Versionshandbuch.
- **Amazon EMR Release 5.32.0** unterstützt Amazon EMR on. EKS Weitere Informationen zu den ersten Schritten mit EMR on EKS finden Sie unter [Was ist Amazon EMR auf EKS](#).
- **Amazon EMR Release 5.32.0** unterstützt Amazon EMR Studio (Preview). Weitere Informationen zu den ersten Schritten mit EMR Studio finden Sie unter [Amazon EMR Studio \(Vorversion\)](#).
- **Verwaltete Richtlinien mit Geltungsbereich:** Um den AWS bewährten Methoden zu entsprechen, EMR hat Amazon EMR verwaltete Standardrichtlinien mit Geltungsbereich v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [EMR Von Amazon verwaltete Richtlinien](#).

Bekannte Probleme

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Websites wie SparkUIs, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von of Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit

der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standardeinstellung eine Ulimit-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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

EMRCluster, auf denen Amazon Linux oder Amazon Linux 2 Amazon Machine Images (AMIs) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen EC2 Amazon-Instances, auf denen das standardmäßige Amazon Linux ausgeführt wird. Wenn neue Amazon Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel und CUDA Updates) NVIDIA, verfügbar werden, nachdem eine EMR Amazon-Version verfügbar wird, laden EMR Cluster-Instances, die standardmäßig ausgeführt werden, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon so anpassen EMR AMI](#), dass es [das neueste Amazon Linux verwendet AMI](#).

- Die Konsolenunterstützung zur Erstellung einer Sicherheitskonfiguration, die die AWS Ranger-Integrationsoption spezifiziert, wird derzeit in der GovCloud Region nicht unterstützt. Die Sicherheitskonfiguration kann mit dem CLI vorgenommen werden. Weitere Informationen finden Sie unter [EMR Sicherheitskonfiguration erstellen](#) im Amazon EMR Management Guide.
- Wenn AtRestEncryption die HDFS Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```

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
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#\$%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.32.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.1.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.17.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMRS3Select-Anschluss
emrfs	2.45.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-0	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.1-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-0	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-0	HTTPEndpunkt für OperationenHDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	2.10.1-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.10.1-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.10.1-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-3	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-3	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.7-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-3	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-3	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.7-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	Mein SQL Datenbankserver.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit

Komponente	Version	Beschreibung
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4,14,3- -1,4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.240.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.240.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.240.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.7-amzn-0	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.4.7-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Clients.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.32.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon leitet EMR nur Rekonfigurationsaktionen für die Klassifizierungen ein, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

emr-5.32.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei container-executor.cfg von HadoopYARN.	Not available.
container-log4j	Ändern Sie die Werte in der Datei container-log4j.properties YARN von Hadoop.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in 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	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	EMRFS-Einstellungen ändern.	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	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	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	Ändert die Werte in der log4j.properties-Datei in 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	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Amazon EMR — kuratierte Einstellungen für Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Werte in der Umgebung ändern HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.	Not available.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Werte in HCatalog der Umgebung ändern.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log 4j2.properties von Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Werte in der Umgebung ändern. HTTPFS	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS	Not available.
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS	Not available.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.	Restarts EMR record server.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-orahome-site	Ändern Sie die Werte in der Datei orahome-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Werte in der YARN Umgebung ändern.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

EMR Amazon-Version 5.31.1

5.31.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
AWS SDK for 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	-	-	-	-
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

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,238,3	0,238,3	0.232	0.232
Spark	2.4.6	2.4.6	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
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

5.31.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPCE S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-`

`$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys.](#)

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.31.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-

Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.0	Amazon SageMaker Spark SDK
emr-ddb	4.15.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.15.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMRS3Select-Anschluss
emrfs	2.43.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.0-amzn-0.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.0-amzn-0.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.0-amzn-0.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.0-amzn-0.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.0-amzn-0.1	HTTPEndpunkt für OperationenHDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	2.10.0-amzn-0.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.10.0-amzn-0.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.0-amzn-0.1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-1	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.7-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.7-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.238.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.238.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.6-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.6-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.6-amzn-0.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.6-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.31.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.31.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS

Klassifizierungen	Beschreibung
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.31.0

5.31.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
AWS SDK for Java	1.11.852	1,11.759	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	-	-	-	-
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

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
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	0,238,3	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

5.31.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.31.0. Änderungen beziehen sich auf Version 5.30.1.

Erste Version: 09. Oktober 2020

Letzte Aktualisierung: 15. Oktober 2020

Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.13.0 aktualisiert
- Amazon SageMaker Spark wurde SDK auf Version 1.4.0 aktualisiert
- Der Amazon-Kinesis-Konnektor wurde auf Version 3.5.9 aktualisiert

- Auf Version AWS SDK for Java 1.11.852 aktualisiert
- BigTop-Tomcat wurde auf Version 8.5.56 aktualisiert
- EMRFS wurde auf Version 2.43.0 aktualisiert
- EMR MetricsAndEventsApiGateway Der Client wurde auf Version 1.4.0 aktualisiert
- EMRS3 Dist CP wurde auf Version 2.15.0 aktualisiert
- EMRS3 Select wurde auf Version 1.6.0 aktualisiert
- Flink wurde auf Version 1.11.0 aktualisiert
- Hadoop wurde auf Version 2.10.0 aktualisiert
- Hive wurde auf Version 2.3.7 aktualisiert
- Hudi wurde auf Version 0.6.0 aktualisiert
- Hue wurde auf Version 4.7.1 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Mxnet wurde auf Version 1.6.0 aktualisiert
- OpenCV wurde auf Version 4.3.0 aktualisiert
- Presto wurde auf Version 0.238.3 aktualisiert
- Auf Version TensorFlow 2.1.0 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.

- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- [Hive-Spaltenstatistiken](#) werden für EMR Amazon-Versionen 5.31.0 und höher unterstützt.
- Aktualisierte Komponentenversionen.
- EMRFSS3EC V2-Unterstützung in Amazon EMR 5.31.0. In den SDK S3-Java-Versionen 1.11.837 und höher wurde der Verschlüsselungsclient Version 2 (S3EC V2) mit verschiedenen Sicherheitsverbesserungen eingeführt. Weitere Informationen finden Sie hier:
 - S3 Blogbeitrag: [Updates für den Amazon S3 Encryption Client](#).
 - AWS SDK for Java Entwicklerhandbuch: [Migrieren Sie Verschlüsselungs- und Entschlüsselungsclients auf V2](#).
 - EMRVerwaltungshandbuch: [Clientseitige Amazon S3 S3-Verschlüsselung](#).

Der Encryption Client V1 ist aus SDK Gründen der Abwärtskompatibilität weiterhin in der verfügbar.

Neue Features

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMRAmazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von ofAmazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit

einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine Ulimit-Standard-Einstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- Mit Amazon EMR 5.31.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration ermöglicht eine detaillierte Datenfilterung auf Spaltenebene für Datenbanken und Tabellen im Glue-Datenkatalog. AWS ermöglicht auch föderiertes Single Sign-On für EMR Notebooks oder Apache Zeppelin von einem Unternehmensidentitätssystem aus. Weitere Informationen finden Sie unter [Integration von Amazon EMR mit AWS Lake Formation](#) im Amazon EMR Management Guide.

Amazon EMR with Lake Formation ist derzeit in 16 AWS Regionen verfügbar: USA Ost (Ohio und Nord-Virginia), USA West (Nordkalifornien und Oregon), Asien-Pazifik (Mumbai, Seoul, Singapur, Sydney und Tokio), Kanada (Zentral), Europa (Frankfurt, Irland, London, Paris und Stockholm), Südamerika (São Paulo).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn AtRestEncryption die HDFS Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
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
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie.

Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.31.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.15,0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.15.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMRS3Select-Anschluss
emrfs	2.43.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.0	Verwaltung von Ressourcen auf EMR Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.0-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.10.0-amzn-0	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.0-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.10.0-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.10.0-amzn-0	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.10.0-amzn-0	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	2.10.0-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.10.0-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.10.0-amzn-0	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.0-amzn-0	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.10.0-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-1	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.

Komponente	Version	Beschreibung
hive-client	2.3.7-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.7-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.238.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.238.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.6-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.6-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.6-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.6-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.31.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.31.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN container-log4j.properties von Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>hive-exec-log4j2.properties</code> von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>llap-daemon-log4j2.properties</code> von Hive.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.

Klassifizierungen	Beschreibung
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.30.2

5.30.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
AWS SDK for Java	1.11.759	1,11.759	1,11.759	1,11,682
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

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
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
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

5.30.2 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPCE S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-`

`$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys.](#)

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.30.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-

Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMRS3Select-Anschluss
emrfs	2.40.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.10.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-6.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-6.1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-6.1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-6.1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-6.1	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	2.8.5-amzn-6.1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-6.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.6-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle zu hcatalog bereitstellt.
hive-client	2.3.6-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.6-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.6.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.5-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.5-amzn-0.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.5-amzn-0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.30.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.30.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.30.1

5.30.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
AWS SDK for 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

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
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
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.227	0.227
Spark	2.4.5	2.4.5	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

5.30.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.30.1. Änderungen beziehen sich auf Version 5.30.0.

Erstveröffentlichungsdatum: 30. Juni 2020

Letzte Aktualisierung: 24. August 2020

Änderungen, Verbesserungen und behobene Probleme

- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Es wurde ein Problem behoben, bei dem der Instance-Controller-Prozess unendlich viele Prozesse ausgelöst hat.
- Es wurde ein Problem behoben, bei dem Hue keine Hive-Abfrage ausführen konnte, die Meldung „Datenbank ist gesperrt“ angezeigt wurde und die Ausführung von Abfragen verhindert wurde.
- Ein Spark-Problem wurde behoben, sodass mehr Aufgaben gleichzeitig auf dem Cluster ausgeführt werden konnten. EMR
- Es wurde ein Problem mit dem Jupyter Notebook behoben, das den Fehler „Zu viele Dateien öffnen“ auf dem Jupyter-Server verursachte.
- Ein Problem mit den Startzeiten von Clustern wurde behoben.

Neue Features

- Die persistenten Anwendungsschnittstellen der Tez-Benutzeroberfläche und des YARN Timeline-Servers sind in den EMR Amazon-Versionen 6.x und 5.30.1 und EMR höher verfügbar. Durch den Linkzugriff mit einem Klick auf den persistenten Anwendungsverlauf können Sie schnell auf den Jobverlauf zugreifen, ohne einen Web-Proxy über eine Verbindung einrichten zu müssen. SSH Protokolle für aktive und beendete Cluster sind 30 Tage nach Ende der Anwendung verfügbar. Weitere Informationen finden Sie unter [Benutzerschnittstellen für persistente Anwendungen anzeigen](#) im Amazon EMR Management Guide.
- EMR Notebook Execution APIs ist verfügbar, um EMR Notebooks über ein Skript oder eine Befehlszeile auszuführen. Durch die Möglichkeit, EMR Notebook-Ausführungen ohne die AWS Konsole zu starten, zu beenden, aufzulisten und zu beschreiben, können Sie ein Notebook programmgesteuert steuern. EMR Mithilfe einer parametrisierten Notebook-Zelle können Sie verschiedene Parameterwerte an ein Notebook übergeben, ohne für jeden neuen Satz von Parameterwerten eine Kopie des Notebooks erstellen zu müssen. [Siehe Aktionen. EMR API](#) Beispielcode finden Sie unter [Beispielbefehle zur programmgesteuerten Ausführung von EMR Notebooks](#).

Bekannte Probleme

- Niedrigerer Grenzwert für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMRAmazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von ofAmazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-StandardEinstellung eine Ulimit-StandardEinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgebung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- EMRNotizbücher

Die Funktion, mit der Sie Kernel und zusätzliche Python-Bibliotheken auf dem primären Clusterknoten installieren können, ist in EMR Version 5.30.1 standardmäßig deaktiviert. Weitere Informationen zu diesem Feature finden Sie unter [Installation von Kernels und Python-Bibliotheken auf einem Cluster-Primärknoten](#).

Um das Feature zu aktivieren, gehen Sie wie folgt vor:

1. Stellen Sie sicher, dass die der Servicerolle für EMR Notebooks zugeordnete Berechtigungsrichtlinie die folgende Aktion zulässt:

```
elasticmapreduce:ListSteps
```

Weitere Informationen finden Sie unter [Servicerolle für EMR Notebooks](#).

2. Verwenden Sie den AWS CLI , um einen Schritt auf dem Cluster auszuführen, der EMR Notebooks einrichtet, wie im folgenden Beispiel gezeigt. Ersetzen *us-east-1* mit der Region, in der sich Ihr Cluster befindet. Weitere Informationen finden Sie unter [Hinzufügen von Schritten zu einem Cluster mithilfe 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"]
```

- Verwaltete Skalierung

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den EMR Amazon-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Job Presto nicht benötigt.

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U

+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.30.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMRS3Select-Anschluss
emrfs	2.40.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.10.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-6	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-https-server	2.8.5-amzn-6	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	2.8.5-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.8.5-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.

Komponente	Version	Beschreibung
hive-client	2.3.6-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.6-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.6.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.5-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.5-amzn-0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.5-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.30.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.30.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN container-log4j.properties von Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>hive-exec-log4j2.properties</code> von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>llap-daemon-log4j2.properties</code> von Hive.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.

Klassifizierungen	Beschreibung
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.30.0

5.30.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
AWS SDK for 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

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
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
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

5.30.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.30.0. Änderungen beziehen sich auf Version 5.29.0.

Erste Version: 13. Mai 2020

Letzte Aktualisierung: 25. Juni 2020

Upgrades

- Auf Version AWS SDK for Java 1.11.759 aktualisiert
- Amazon SageMaker Spark wurde SDK auf Version 1.3.0 aktualisiert
- EMRRecord Server wurde auf Version 1.6.0 aktualisiert
- Flink auf Version 1.10.0 aktualisiert
- Ganglia auf Version 3.7.2 aktualisiert
- Auf Version HBase 1.4.13 aktualisiert
- Hudi auf Version 0.5.2-incubating aktualisiert
- Hue auf Version 4.6.0 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Livy auf Version 0.7.0-incubating aktualisiert
- Oozie auf Version 5.2.0 aktualisiert
- Presto auf Version 0.232 aktualisiert
- Spark auf Version 2.4.5 aktualisiert
- Aktualisierte Konnektoren und Treiber: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; DynamoDB Connector 4.14.0 EMR

Neue Features

- EMRNotebooks — Bei Verwendung mit EMR Clustern, die mit 5.30.0 erstellt wurden, werden Notebook-Kernel auf dem Cluster ausgeführt. EMR Dies verbessert die Notebook-Leistung und ermöglicht es Ihnen, Kernel zu installieren und anzupassen. Sie können Python-Bibliotheken auch auf dem Cluster-Primärknoten installieren. Weitere Informationen finden Sie unter [Installation und Verwendung von Kernels and Libraries im Management Guide](#). EMR
- Verwaltung der Skalierung — Mit EMR Amazon-Version 5.30.0 und höher können Sie EMR Managed Scaling aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster je nach Arbeitslast

automatisch zu erhöhen oder zu verringern. Amazon wertet EMR kontinuierlich Cluster-Metriken aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster im Hinblick auf Kosten und Geschwindigkeit optimieren. Weitere Informationen finden Sie unter [Scaling Cluster Resources](#) im Amazon EMR Management Guide.

- In Amazon S3 gespeicherte Protokolldateien verschlüsseln — Mit Amazon EMR Version 5.30.0 und höher können Sie in Amazon S3 gespeicherte Protokolldateien mit einem AWS KMS vom Kunden verwalteten Schlüssel verschlüsseln. Weitere Informationen finden Sie unter [Verschlüsseln von in Amazon S3 gespeicherten Protokolldateien](#) im Amazon EMR Management Guide.
- Amazon Linux 2-Unterstützung — In EMR Version 5.30.0 und höher, EMR uses Amazon Linux 2 OS. Das neue benutzerdefinierte AMIs (Amazon Machine Image) muss auf the Amazon Linux 2 basieren AMI. Weitere Informationen finden Sie unter [Benutzerdefiniert verwenden AMI](#).
- Presto Graceful Auto Scale — Für EMR Cluster, die 5.30.0 verwenden, kann ein Timeout für die automatische Skalierung festgelegt werden, sodass Presto-Aufgaben Zeit haben, ihre Ausführung zu beenden, bevor ihr Knoten außer Betrieb genommen wird. Weitere Informationen finden Sie unter [Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden](#).
- Erstellung einer Flotteninstanz mit neuer Zuweisungsstrategieoption — Eine neue Zuweisungsstrategieoption ist in Version 5.12.1 und höher verfügbar. EMR Sie bietet eine schnellere Cluster-Bereitstellung, eine genauere Spot-Zuweisung und weniger Unterbrechungen von Spot Instances. Aktualisierungen für nicht standardmäßige EMR Servicerollen sind erforderlich. Sehen Sie unter [Konfigurieren von Instance-Flotten](#).
- Befehle `sudo systemctl stop` und `sudo systemctl start` — In EMR Version 5.30.0 und höher, die das Betriebssystem use Amazon Linux 2 verwendet, und Befehle, um Dienste neu zu starten. EMR `sudo systemctl stop` `sudo systemctl start` Weitere Informationen finden Sie unter [Wie starte ich einen Service bei Amazon neu EMR?](#) .

Änderungen, Verbesserungen und behobene Probleme

- EMR Version 5.30.0 installiert Ganglia nicht standardmäßig. Sie können Ganglia explizit für die Installation auswählen, wenn Sie einen Cluster erstellen.
- Spark-Leistungsoptimierungen.
- Presto-Leistungsoptimierungen.
- Python 3 ist der Standard für EMR Amazon-Version 5.30.0 und höher.
- Die standardmäßige verwaltete Sicherheitsgruppe für den Zugriff auf Services in privaten Subnetzen wurde mit neuen Regeln aktualisiert. Wenn Sie benutzerdefinierte Sicherheitsgruppe für den Servicezugriff verwenden, müssen Sie dieselben Regeln wie die standardmäßige

verwaltete Sicherheitsgruppe einschließen. Weitere Informationen finden Sie unter [Amazon EMR - Managed Security Group for Service Access \(Private Subnets\)](#). Wenn Sie eine benutzerdefinierte Servicerolle für Amazon verwenden, müssen Sie die entsprechenden Berechtigungen erteilen, damit `ec2:describeSecurityGroups` überprüft werden kann, ob die Sicherheitsgruppen korrekt erstellt wurden. Wenn Sie `EMR_DefaultRole` verwenden, ist diese Berechtigung bereits in der standardmäßigen verwalteten Richtlinie enthalten.

Bekannte Probleme

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen `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` und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standardeinstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein `ulimit` über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- **Verwaltete Skalierung**

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den EMR Amazon-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Job Presto nicht benötigt.

- **Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung**

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Die Standard-Datenbank-Engine für Hue 4.6.0 ist SQLite, was zu Problemen führt, wenn Sie versuchen, Hue mit einer externen Datenbank zu verwenden. Um dieses Problem zu beheben, setzen Sie `engine` in Ihrer `hue.ini` Konfigurationsklassifizierung auf `mysql`. Dieses Problem wurde in EMR Amazon-Version 5.30.1 behoben.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.

- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

5.30.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.0.0	Conda env für emr Notebook
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMRS3Select-Anschluss
emrfs	2.40.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.10.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-6	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-6	HTTP-Endpunkt für Operationen in HDFS.
hadoop-kms-server	2.8.5-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.8.5-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.13	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.13	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.13	HBase Befehlszeilenclient.
hbase-rest-server	1.4.13	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.13	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.6-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.6-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.6-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.6.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.5-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.5-amzn-0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.5-amzn-0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.30.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.30.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4 j-yarn-session .properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.29.0

5.29.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
AWS SDK for 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
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

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
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	-	-	-	-
Jupyter Enterprise Gateway	-	-	-	-
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
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

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
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

5.29.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.29.0. Änderungen beziehen sich auf Version 5.28.1.

Erste Version: 17. Januar 2020

Upgrades

- Auf Version AWS SDK for Java 1.11.682 aktualisiert
- Upgrade von Hive auf Version 2.3.6
- Flink wurde auf Version 1.9.1 aktualisiert
- EmrFS wurde auf Version 2.38.0 aktualisiert
- EMRDynamoDB Connector wurde auf Version 4.13.0 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Spark-Leistungsoptimierungen.
- EMRFS
 - Management Guide aktualisiert die Standardeinstellungen von emrfs-site.xml für eine konsistente Ansicht.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.29.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.13.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.12.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.4.0	EMRS3Select-Anschluss
emrfs	2.38.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.9.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-5	HTTP-Endpunkt für Operationen in HDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.5-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.10	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.10	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.10	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.10	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.10	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.6-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.6-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.6-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.5.0-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.227	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.227	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.227	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)

Komponente	Version	Beschreibung
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführung engine für YARN
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.29.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.29.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS

Klassifizierungen	Beschreibung
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.28.1

5.28.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
AWS SDK for 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
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

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
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
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

5.28.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.28.1. Änderungen beziehen sich auf Version 5.28.0.

Erste Version: 10. Januar 2020

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Spark-Kompatibilitätsprobleme behoben.
- CloudWatch Metriken

- Fehler bei der Veröffentlichung von Amazon CloudWatch Metrics auf einem EMR Cluster mit mehreren Primärknoten behoben.
- Deaktivierte Protokollnachricht
 - Falsche Protokollnachricht „... verwendet die alte Version (<4.5.8) des Apache-HTTP-Clients“ deaktiviert.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.28.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.37.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.9.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-5	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.8.5-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.10	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.10	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.10	HBase Befehlszeilenclient.
hbase-rest-server	1.4.10	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.10	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.6-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	2.3.6-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.0-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.227	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.227	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.227	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.28.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.28.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS

Klassifizierungen	Beschreibung
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.28.0

5.28.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
AWS SDK for 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	-	-	-	-
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

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
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

5.28.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.28.0. Änderungen beziehen sich auf Version 5.27.0.

Erste Version: 12. November 2019

Upgrades

- Upgrade von Flink auf Version 1.9.0
- Upgrade von Hive auf Version 2.3.6
- Auf Version MXNet 1.5.1 aktualisiert
- Upgrade von Phoenix auf Version 4.14.3

- Upgrade von Presto auf Version 0.227
- Upgrade von Zeppelin auf Version 0.8.2

Neue Features

- [Apache Hudi](#) kann jetzt von Amazon installiert EMR werden, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Hudi](#).
- (25. November 2019) Sie können jetzt auf Wunsch mehrere Schritte gleichzeitig ausführen, um die Cluster-Nutzung zu verbessern und Kosten zu sparen. Außerdem können Sie sowohl anstehende als auch laufende Schritte stornieren. Weitere Informationen finden Sie unter [Arbeiten mit Schritten mithilfe der AWS CLI AND-Konsole](#).
- (3. Dezember 2019) Sie können jetzt EMR Cluster auf erstellen und ausführen AWS Outposts. AWS Outposts ermöglicht native AWS Dienste, Infrastrukturen und Betriebsmodelle in lokalen Einrichtungen. In AWS Outposts Umgebungen können Sie dieselben AWS APIs Tools und dieselbe Infrastruktur verwenden wie in der AWS Cloud. Weitere Informationen finden Sie unter [EMRCluster on AWS Outposts](#).
- (11. März 2020) Ab EMR Amazon-Version 5.28.0 können Sie EMR Amazon-Cluster in einem Local Zones-Subnetz als logische Erweiterung einer AWS Region, die AWS Local Zones unterstützt, erstellen und ausführen. Eine lokale Zone ermöglicht es, EMR Amazon-Funktionen und eine Untergruppe von AWS Diensten, wie Rechen- und Speicherdienste, näher an den Benutzern zu platzieren, wodurch der Zugriff auf lokal ausgeführte Anwendungen mit sehr geringer Latenz ermöglicht wird. Eine Liste der verfügbaren Local Zones finden Sie unter [AWS Local Zones](#). Informationen zum Zugriff auf verfügbare AWS Local Zones finden Sie unter [Regionen, Availability Zones und Local Zones](#).

Local Zones unterstützen derzeit keine EMR Amazon-Notebooks und unterstützen keine direkten Verbindungen zu Amazon EMR über den VPC Schnittstellenendpunkt (AWS PrivateLink).

Änderungen, Verbesserungen und behobene Probleme

- Erweiterter Anwendungssupport für Cluster mit hoher Verfügbarkeit
 - Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR Cluster mit mehreren Primärknoten](#) im Amazon EMR Management Guide.
- Spark
 - Leistungsoptimierungen

- Hive
 - Leistungsoptimierungen
- Presto
 - Leistungsoptimierungen

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.28.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.37.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.9.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-5	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.8.5-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.10	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.10	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.10	HBase Befehlszeilenclient.
hbase-rest-server	1.4.10	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.10	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.6-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	2.3.6-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.0-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,3- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.227	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.227	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.227	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.28.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.28.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS

Klassifizierungen	Beschreibung
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.27.1

5.27.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
AWS SDK for Java	1.11.615	1,11,615	1,11.595	1,11.566
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

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1
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

5.27.1 Versionshinweise

Dies ist eine Patch-Version. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Unterstützungsstatus von Instance Metadata Service (IMDS) V2: Die Komponenten Amazon EMR 5.23.1, 5.27.1 und 5.32 oder höher werden für alle Aufrufe verwendet. IMDSv2 IMDS Für IMDS Aufrufe in Ihrem Anwendungscode können Sie sowohl als auch IMDSv1 verwenden oder die Option so konfigurierenIMDSv2, dass sie nur für zusätzliche Sicherheit IMDS verwendet wird. IMDSv2 Bei anderen EMR 5.x-Versionen IMDSv1 führt die Deaktivierung zu einem Fehler beim Starten des Clusters.

5.27.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.36.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-4	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.10	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.10	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.10	HBase Befehlszeilenclient.
hbase-rest-server	1.4.10	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.10	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.5-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.5-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	2.3.5-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	4.14.2- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,2- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.224	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.224	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungsengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.27.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.27.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.27.0

5.27.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
AWS SDK for 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

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
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
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

5.27.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.27.0. Änderungen beziehen sich auf Version 5.26.0.

Erste Version: 23. September 2019

Upgrades

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1,14.0
- Konnektoren und Treiber:
 - DynamoDB Connector 4.12.0

Neue Features

- (24. Oktober 2019) Die folgenden neuen Funktionen in EMR Notizbüchern sind in allen EMR Amazon-Versionen verfügbar.
 - Sie können jetzt Git-Repositorys mit EMR Notizbüchern verknüpfen, um Ihre Notizbücher in einer versionskontrollierten Umgebung zu speichern. Sie können Code mit Kollegen teilen und vorhandene Jupyter Notebooks über Remote-Git-Repositorys wiederverwenden. Weitere Informationen finden Sie unter [Git-Repositorys mit Amazon EMR Notebooks verknüpfen](#) im Amazon EMR Management Guide.
 - Das [Hilfsprogramm nbdime](#) ist jetzt in EMR Notebooks verfügbar, um das Vergleichen und Zusammenführen von Notizbüchern zu vereinfachen.
 - EMR Notebooks werden jetzt unterstützt. JupyterLab JupyterLab ist eine webbasierte interaktive Entwicklungsumgebung, die vollständig mit Jupyter-Notebooks kompatibel ist. Sie können jetzt wählen, ob Sie Ihr Notizbuch entweder im JupyterLab Jupyter-Notebook-Editor öffnen möchten.
- (30. Oktober 2019) Mit den EMR Amazon-Versionen 5.25.0 und höher können Sie über die Cluster-Übersichtsseite oder den Tab Anwendungsverlauf in der Konsole eine Verbindung zur Benutzeroberfläche des Spark-Verlaufs herstellen. Anstatt einen Web-Proxy über eine SSH Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive

und beendete Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbensutzeroberflächen](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- EMRAmazon-Cluster mit mehreren Primärknoten
 - Sie können Flink auf einem EMR Amazon-Cluster mit mehreren Primärknoten installieren und ausführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen und Features](#).
 - Sie können HDFS transparente Verschlüsselung auf einem EMR Amazon-Cluster mit mehreren Primärknoten konfigurieren. Weitere Informationen finden Sie unter [HDFSTransparente Verschlüsselung auf EMR Clustern mit mehreren Primärknoten](#).
 - Sie können jetzt die Konfiguration von Anwendungen ändern, die auf einem EMR Amazon-Cluster mit mehreren Primärknoten ausgeführt werden. Weitere Informationen finden Sie unter [Angaben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).
- EMRAmazon-DynamoDB-Konnektor
 - Amazon EMR -DynamoDB Connector unterstützt jetzt die folgenden DynamoDB-Datentypen: boolean, list, map, item, null. Weitere Informationen finden Sie unter [Eine Hive-Tabelle zur Ausführung von Hive-Befehlen einrichten](#).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.27.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.3.0	EMRS3Select-Anschluss
emrfs	2.36.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.8.5-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.10	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.10	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.10	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.5-amzn-1	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-1	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.5-amzn-1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.5-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.5-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,2- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.224	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.224	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)

Komponente	Version	Beschreibung
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.27.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.27.0 Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von RangerKMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS

Klassifizierungen	Beschreibung
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.26.0

5.26.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
AWS SDK for 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	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.220	0.219	0.219
Spark	2.4.3	2.4.3	2.4.2	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.13.1	1.12.0	1.12.0
Tez	0.9.2	0.9.2	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.13	3.4.13

5.26.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.26.0. Änderungen beziehen sich auf Version 5.25.0.

Erste Version: 8. August 2019

Letzte Aktualisierung: 19. August 2019

Upgrades

- AWS SDK for Java 1.11.595
- HBase1.4,10
- Phoenix 4.14.2

- Konnektoren und Treiber:
 - DynamoDB Connector 4.11.0
 - MariaDB Connector 2.4.2
 - Amazon Redshift JDBC Redshift-Treiber 1.2.32.1056

Neue Features

- (Beta) Mit Amazon EMR 5.26.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration bietet einen detaillierten Zugriff auf Spaltenebene auf Datenbanken und Tabellen im Glue-Datenkatalog. AWS ermöglicht auch föderiertes Single Sign-On für EMR Notebooks oder Apache Zeppelin von einem Unternehmensidentitätssystem aus. Weitere Informationen finden Sie unter [Integration von Amazon EMR mit AWS Lake Formation \(Beta\)](#).
- (19. August 2019) Amazon EMR Block Public Access ist jetzt in allen EMR Amazon-Versionen verfügbar, die Sicherheitsgruppen unterstützen. Den öffentlichen Zugriff blockieren ist eine kontoweite Einstellung, die für jede AWS Region gilt. Den öffentlichen Zugriff blockieren verhindert, dass ein Cluster gestartet wird, wenn eine dem Cluster zugeordnete Sicherheitsgruppe über eine Regel verfügt, die eingehenden Datenverkehr von IPv4 0.0.0.0/0 oder IPv6:: /0 (öffentlicher Zugriff) auf einem Port zulässt, sofern kein Port als Ausnahme angegeben ist. Port 22 ist standardmäßig eine Ausnahme. Weitere Informationen finden Sie unter [Using Amazon EMR Block Public Access](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- EMR-Notebooks
 - Ab Version EMR 5.26.0 unterstützt EMR Notebooks zusätzlich zu den Standard-Python-Bibliotheken auch Python-Bibliotheken für Notebooks. Sie können Bibliotheken für Notebooks vom Notebook-Editor aus installieren, ohne einen Cluster neu erstellen oder ein Notebook erneut an einen Cluster anhängen zu müssen. Notebook-bezogene Bibliotheken werden in einer virtuellen Python-Umgebung erstellt und gelten daher nur für die aktuelle Notebook-Sitzung. Auf diese Weise können Sie Abhängigkeiten zwischen Notebooks isolieren. Weitere Informationen finden Sie unter [Using Notebook Scoped Libraries](#) im Amazon EMR Management Guide.
- EMRFS
 - Sie können eine ETag Bestätigungsfunktion (Beta) aktivieren, indem Sie auf `einstellenfs.s3.consistent.metadata.etag.verification.enabled.true` Mit dieser Funktion überprüft Amazon EMRFS S3, ob ETags es sich bei den gelesenen Objekten

um die neueste verfügbare Version handelt. Diese Funktion ist hilfreich für read-after-update Anwendungsfälle, in denen Dateien auf Amazon S3 überschrieben werden, während derselbe Name beibehalten wird. Diese ETag Überprüfungsfunktion funktioniert derzeit nicht mit S3 Select. Weitere Informationen finden Sie unter [Konfigurieren der konsistenten Ansicht](#).

- Spark
 - Die folgenden Optimierungen sind jetzt standardmäßig aktiviert: dynamisches Bereinigen von Partitionen DISTINCT zuvor INTERSECT, Verbesserungen bei der Ableitung von SQL Planstatistiken für JOIN nachfolgende DISTINCT Abfragen, Reduzierung skalarer Unterabfragen, optimierte Neureihenfolge von Verknüpfungen und Bloom-Filter-Join. Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).
 - Die Codegenerierung für die gesamte Phase für Sort Merge Join wurde verbessert.
 - Die Wiederverwendung von Abfragefragmenten und Unterabfragen wurde verbessert.
 - Verbesserungen bei der Vorabzuweisung von Executoren beim Start von Spark.
 - Bloom-Filter-Joins werden nicht mehr angewendet, wenn die kleinere Seite der Verknüpfung einen Broadcast-Hinweis enthält.
- Tez
 - Ein Problem mit Tez wurde behoben. Tez UI funktioniert jetzt auf einem EMR Amazon-Cluster mit mehreren Primärknoten.

Bekannte Probleme

- Die verbesserten Funktionen zur Codegenerierung in der gesamten Phase für Sort Merge Join können den Speicherdruck erhöhen, wenn sie aktiviert sind. Diese Optimierung verbessert die Leistung, kann jedoch zu Wiederholungsversuchen oder Fehlschlägen führen, wenn der nicht so eingestellt `spark.yarn.executor.memoryOverheadFactor` ist, dass er genügend Speicher zur Verfügung stellt. Um dieses Feature zu deaktivieren, setzen Sie `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` auf „false“.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite

Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.26.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.11.0	Amazon DynamoDB-Konnektor für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.10.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.12.0	Für Amazon S3 optimierte verteilte Kopieranwendung.
emr-s3-select	1.3.0	EMRS3Select-Anschluss
emrfs	2.35.0	Amazon S3-Konnektor für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.10	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.10	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.10	HBaseBefehlszeilenclient.
hbase-rest-server	1.4.10	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.10	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.5-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.5-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.5-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.5-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,2- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.220	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.220	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.3	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.13.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

5.26.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.26.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.25.0

5.25.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
AWS SDK for 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

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
HBase	1.4.9	1.4.9	1.4.9	1.4.9
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

5.25.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.25.0. Änderungen beziehen sich auf Version 5.24.1.

Erste Version: 17. Juli 2019

Letzte Aktualisierung: 30. Oktober 2019

Amazon EMR 5,25,0

Upgrades

- 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

Neue Features

- (30. Oktober 2019) Ab EMR Amazon-Version 5.25.0 können Sie über die Cluster-Übersichtsseite oder den Tab Anwendungsverlauf in der Konsole eine Verbindung zur Benutzeroberfläche des Spark-Verlaufs herstellen. Anstatt einen Web-Proxy über eine SSH Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete

Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbenuzoberflächen](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Die Leistung einiger Verknüpfungen wurde verbessert, indem Bloom-Filter zum Vorfiltern von Eingaben verwendet wurden. Die Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.bloomFilterJoin.enabled` auf `true` gesetzt wird.
 - Die Leistung bei der Gruppierung nach Spalten vom Typ Zeichenfolge wurde verbessert.
 - Die standardmäßige Speicher- und Kernkonfiguration von Spark Executor für R4-Instance-Typen für Cluster ohne Installation wurde verbessert. HBase
 - Es wurde ein früheres Problem mit dem Feature zum dynamischen Bereinigen von Partitionen behoben, bei dem sich die bereinigte Tabelle auf der linken Seite des Joins befinden musste.
 - DISTINCT vor der INTERSECT Optimierung verbessert und gilt nun auch für weitere Fälle, in denen Aliase involviert sind.
 - Die Ableitung von SQL Planstatistiken für Abfragen, JOIN denen gefolgt von DISTINCT Abfragen, wurde verbessert. Diese Verbesserung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.statsImprovements.enabled` auf `true` gesetzt wird. Diese Optimierung ist für das Feature „Unterscheiden vor Überschneiden“ erforderlich und wird automatisch aktiviert, wenn `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true` gesetzt ist.
 - Optimierte Verbindungsreihenfolge basierend auf Tabellengröße und Filtern. Diese Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` auf `true` gesetzt wird.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- EMRFS
 - Die EMRFS Einstellung, `fs.s3.buckets.create.enabled`, ist jetzt standardmäßig deaktiviert. Bei Tests haben wir festgestellt, dass die Deaktivierung dieser Einstellung die Leistung verbessert und die unbeabsichtigte Erstellung von S3-Buckets verhindert. Wenn Ihre Anwendung auf dieser Funktionalität basiert, können Sie sie aktivieren, indem Sie die Eigenschaft `true` in der Konfigurationsklassifizierung von `fs.s3.buckets.create.enabled`

auf `emrfs-site` setzen. Weitere Informationen finden Sie unter [Angeben einer Konfiguration beim Erstellen eines Clusters](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
 - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
 - Es wurde eine Option hinzugefügt, um die EBS Verschlüsselung mit Version 5.24.0 und höher zu aktivieren. Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten Datenträgers AMI zur Verschlüsselung des Root-Geräte-Volumens.
 - Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#) im Amazon EMR Management Guide.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop` Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.25.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.10.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.9.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.3.0	EMRS3Select-Anschluss
emrfs	2.34.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP-Endpunkt für Operationen auf HDFS.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.9	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.9	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.9	HBaseBefehlszeilenclient.
hbase-rest-server	1.4.9	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.9	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.5-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.

Komponente	Version	Beschreibung
hive-client	2.3.5-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.5-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,1- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.220	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.220	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.3	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.3	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.13.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

5.25.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.25.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN <code>container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xml HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
recordserver-env	Werte in der EMR RecordServer Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in EMR RecordServer der Datei server.properties.
recordserver-log4j	Ändern Sie die Werte in der Datei EMR RecordServer log4j.properties.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.24.1

5.24.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
AWS SDK for 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
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	-	-	-	-

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
JupyterHub	0.9.6	0.9.6	0.9.4	0.9.4
Livy	0.6.0	0.6.0	0.5.0	0.5.0
MXNet	1.4.0	1.4.0	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.219	0.215	0.215
Spark	2.4.2	2.4.2	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
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

5.24.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.24.1. Änderungen beziehen sich auf Version 5.24.0.

Erste Version: 26. Juni 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.24.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.9.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.8.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.33.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-4	HTTP-Endpoint für Operationen mit HDFS.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.9	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.9	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.9	HBase Befehlszeilenclient.
hbase-rest-server	1.4.9	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.9	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.4-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	2.3.4-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	4.14.1- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,1- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.219	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.219	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.2	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.4.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.24.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.24.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.24.0

5.24.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
AWS SDK for 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

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
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	-	-	-	-
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

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
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

5.24.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.24.0. Änderungen beziehen sich auf Version 5.23.0.

Erste Version: 11. Juni 2019

Letzte Aktualisierung: 5. August 2019

Upgrades

- 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

- Konnektoren und Treiber:
 - DynamoDB Connector 4.9.0
 - MariaDB Connector 2.4.1
 - Amazon Redshift JDBC Redshift-Treiber 1.2.27.1051

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Es wurde eine Optimierung für das dynamische Bereinigen von Partitionen hinzugefügt. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.dynamicPartitionPruning.enabled` auf `true`.
 - Die Leistung von INTERSECT-Anfragen wurde verbessert. Diese Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true`.
 - Es wurde eine Optimierung hinzugefügt, um skalare Unterabfragen mit Aggregaten, die dieselbe Beziehung verwenden, zu vereinfachen. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` auf `true`.
 - Die Codegenerierung für die gesamte Phase wurde verbessert.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
 - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
 - Es wurde eine Option hinzugefügt, um die Verschlüsselung zu aktivieren. EBS Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten `VolumesAMI`, um das Root-Geräte-Volumen zu verschlüsseln.
 - Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#) im Amazon EMR Management Guide.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.24.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.9.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.8.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.3.0	EMRS3Select-Anschluss
emrfs	2.33.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPEndpunkt für OperationenHDFS.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.9	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.9	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.9	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.9	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.9	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-2	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.4-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,14,1- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.219	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.219	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.24.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.24.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN container-log4j.properties von Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_conf.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.23.1

5.23.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
AWS SDK for 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

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
HBase	1.4.9	1.4.9	1.4.9	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
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

5.23.1 Versionshinweise

Dies ist eine Patch-Version. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Unterstützungsstatus von Instance Metadata Service (IMDS) V2: Die Komponenten Amazon EMR 5.23.1, 5.27.1 und 5.32 oder höher werden für alle Aufrufe verwendet. IMDSv2 IMDS Für IMDS Aufrufe in Ihrem Anwendungscode können Sie sowohl als auch IMDSv1 verwenden oder die Option so konfigurieren IMDSv2, dass sie nur für zusätzliche Sicherheit IMDS verwendet wird. IMDSv2 Bei anderen EMR 5.x-Versionen IMDSv1 führt die Deaktivierung zu einem Fehler beim Starten des Clusters.

5.23.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.8.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.7.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.32.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-3	HTTP-Endpoint für Operationen auf HDFS.
hadoop-kms-server	2.8.5-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.9	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.9	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.9	HBase Befehlszeilenclient.
hbase-rest-server	1.4.9	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.9	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.4-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-1	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle zu HCatalog bereitstellt.
hive-client	2.3.4-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,1- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Cli-ent.

5.23.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.23.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.23.0

5.23.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
AWS SDK for 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
Jupyter Enterprise Gateway	-	-	-	-
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

5.23.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.23.0. Änderungen beziehen sich auf Version 5.22.0.

Erste Version: 01. April 2019

Letzte Aktualisierung: 30. April 2019

Upgrades

- AWS SDK for Java 1.11.519

Neue Features

- (30. April 2019) Mit Amazon EMR 5.23.0 und höher können Sie einen Cluster mit drei primären Knoten starten, um die Hochverfügbarkeit von Anwendungen wie YARN Resource Manager, Spark HDFS NameNode, Hive und Ganglia zu unterstützen. Der Primärknoten ist mit diesem Feature keine potenzielle einzelne Fehlerquelle mehr. Wenn einer der Primärknoten ausfällt, wechselt Amazon EMR automatisch zu einem Standby-Primärknoten und ersetzt den ausgefallenen Primärknoten durch einen neuen mit derselben Konfiguration und denselben Bootstrap-Aktionen. Weitere Informationen finden Sie unter [Primärknoten planen und konfigurieren](#).

Bekannte Probleme

- Tez UI (In EMR Amazon-Version 5.26.0 behoben)

Die Tez-Benutzeroberfläche funktioniert nicht auf einem EMR Cluster mit mehreren Primärknoten.

- Farbton (In EMR Amazon-Version 5.24.0 behoben)

- Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

- Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

- Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop` Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.23.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.8.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.7.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.2.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.32.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFSDienst für die Verwaltung des Hadoop-Dateisystem-Journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-3	HTTPEndpunkt für OperationenHDFS.
hadoop-kms-server	2.8.5-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.9	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.9	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.9	HBase Befehlszeilenclient.
hbase-rest-server	1.4.9	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.9	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.4-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server2	2.3.4-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	4.14.1- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,1- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.23.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.23.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.22.0

5.22.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
AWS SDK for 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

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.0	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.22.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.22.0. Änderungen beziehen sich auf Version 5.21.0.

Important

Ab EMR Amazon-Version 5.22.0 EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Frühere EMR Amazon-Versionen verwenden in einigen Fällen AWS Signature Version 2, sofern in den Versionshinweisen nicht angegeben ist, dass ausschließlich Signature Version 4 verwendet wird. Weitere Informationen finden Sie unter [Authentifizieren von Anfragen \(AWS Signature Version 4\)](#) und [Authentifizieren von Anfragen \(AWS Signature Version 2\)](#) im Amazon Simple Storage Service Developer Guide.

Erste Version: 20. März 2019

Upgrades

- Flink 1.7.1

- HBase 1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Konnektoren und Treiber:
 - DynamoDB Connector 4.8.0
 - MariaDB Connector 2.2.6
 - Amazon Redshift JDBC Redshift-Treiber 1.2.20.1043

Neue Features

- Die EBS Standardkonfiguration für EC2 Instance-Typen mit „-only“-Speicher wurde geändert. EBS Wenn Sie einen Cluster mit Amazon EMR Version 5.22.0 und höher erstellen, erhöht sich die EBS Standardspeichermenge je nach Größe der Instance. Darüber hinaus haben wir den erhöhten Speicherplatz auf mehrere Volumes aufgeteilt, um die Leistung zu erhöhen IOPS. Wenn Sie eine andere EBS Instance-Speicherkonfiguration verwenden möchten, können Sie diese angeben, wenn Sie einen EMR Cluster erstellen oder Knoten zu einem vorhandenen Cluster hinzufügen. Weitere Informationen zur Speichermenge und Anzahl der standardmäßig zugewiesenen Volumes für jeden Instance-Typ finden Sie unter [EBS Standardspeicher für Instances](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Es wurde eine neue Konfigurationseigenschaft für Spark auf YARN, eingeführt `spark.yarn.executor.memoryOverheadFactor`. Der Wert dieser Eigenschaft ist ein Skalierungsfaktor, der den Wert des Speicher-Overheads auf einen Prozentsatz des Executor-Speichers festlegt, mit einem Minimum von 384 MB. Wenn der Speicher-Overhead explizit mit `spark.yarn.executor.memoryOverhead` festgelegt wird, hat diese Eigenschaft keine Auswirkung. Der Standardwert ist `0.1875`, was 18,75 % entspricht. Dieser Standard für Amazon EMR lässt mehr Platz in YARN Containern für den Executor-Speicheraufwand übrig als der intern von Spark festgelegte Standard von 10%. Der EMR Amazon-Standard von 18,75% ergab empirisch weniger speicherbedingte Ausfälle in -DS-Benchmarks. TPC
 - [-26316 wurde zurückportiert, um die Leistung zu verbessern. SPARK](#)

- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS

Bekannte Probleme

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
 - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-

Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.22.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-

Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.8.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.6.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.31.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-2	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.5-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.5-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.5-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.9	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.9	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.9	HBase Befehlszeilenclient.
hbase-rest-server	1.4.9	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.9	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.4-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.4-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,14,1- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Cli-ent.

5.22.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.22.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.21.2

5.21.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
AWS SDK for Java	1.11.479	1.11.479	1.11.479	1.11.461
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	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	-	-	-	-

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Jupyter Enterprise Gateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
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

5.21.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.21.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.30.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.5-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.8	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.8	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.8	HBase Befehlszeilenclient.
hbase-rest-server	1.4.8	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.8	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.21.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.21.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN <code>container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.21.1

5.21.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
AWS SDK for 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
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

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
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	0.215	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

5.21.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.21.1. Änderungen beziehen sich auf Version 5.21.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.21.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.30.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.8	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.8	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.8	HBase Befehlszeilenclient.
hbase-rest-server	1.4.8	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.8	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu hcatalog bereitstellt.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Cli-ent.

5.21.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.21.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.21.0

5.21.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
AWS SDK for 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
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	-	-	-	-

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
Jupyter Enterprise Gateway	-	-	-	-
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
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

5.21.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.21.0. Änderungen beziehen sich auf Version 5.20.0.

Veröffentlichungsdatum: 18. Februar 2019

Letzte Aktualisierung: 3. April 2019

Upgrades

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

Neue Features

- (3. April 2019) Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Sie tun dies, indem Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Änderungen, Verbesserungen und behobene Probleme

- Zeppelin
 - Rückportiert [ZEPPELIN-3878](#).

Bekannte Probleme

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
 - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez

- Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie unter `http://` eine Verbindung zur Tez-Benutzeroberfläche herstellen `MasterDNS:8080/tez-ui` über eine SSH Verbindung zum primären Clusterknoten wird der Fehler „Adaptervorgang fehlgeschlagen - Timeline-Server () ATS ist außer Reichweite“ angezeigt. Entweder ist er ausgefallen oder nicht aktiviert“ CORS wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server sendet und `localhost` nicht den Hostnamen des primären Knotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde.

Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.21.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.30.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.8	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.8	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.8	HBase Befehlszeilenclient.
hbase-rest-server	1.4.8	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.8	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.4-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Cli-ent.

5.21.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.21.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.20.1

5.20.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
AWS SDK for 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
HCatalog	2.3.4	2.3.4	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.2.0	4.2.0
Iceberg	-	-	-	-

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
Jupyter Enterprise Gateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.0	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.214	0.212	0.212
Spark	2.4.0	2.4.0	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.11.0	1.11.0
Tez	0.9.1	0.9.1	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

5.20.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.20.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.29.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.5-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.8	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.8	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.8	HBase Befehlszeilenclient.
hbase-rest-server	1.4.8	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.8	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.214	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.214	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.20.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.20.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.20.0

5.20.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
AWS SDK for 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	-	-	-	-
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

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
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	0.214	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

5.20.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.20.0. Änderungen beziehen sich auf Version 5.19.0.

Veröffentlichungsdatum: 18. Dezember 2018

Letzte Aktualisierung: 22. Januar 2019

Upgrades

- 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

Neue Features

- (22. Januar 2019) Kerberos in Amazon EMR wurde verbessert und unterstützt nun die Authentifizierung von Principals von einem externen System aus. KDC Dadurch wird die Prinzipalverwaltung zentralisiert, da sich mehrere Cluster einen einzigen externen Cluster teilen können. KDC Darüber hinaus KDC kann das externe System eine bereichsübergreifende Vertrauensstellung mit einer Active Directory-Domäne haben. Auf diese Weise können alle Cluster Prinzipale von Active Directory authentifizieren. Weitere Informationen finden Sie unter [Verwenden der Kerberos-Authentifizierung](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Standard Amazon Linux AMI für Amazon EMR
 - Das Python3-Paket wurde von Python 3.4 auf 3.6 aktualisiert.
- Der EMRFS S3-optimierte Committer
 - Der EMRFS S3-optimierte Committer ist jetzt standardmäßig aktiviert, was die Schreibleistung verbessert. Weitere Informationen finden Sie unter [Verwenden Sie den EMRFS S3-optimierten Committer](#).
- Hive
 - [Rückportiert -16686. HIVE](#)
- Mit Spark und Hive Glue
 - In Version EMR 5.20.0 oder höher wird die parallel Partitionsbereinigung automatisch für Spark und Hive aktiviert, wenn AWS Glue Data Catalog als Metastore verwendet wird. Diese Änderung reduziert die Zeit für die Abfrageplanung erheblich, da mehrere Anfragen parallel ausgeführt werden, um Partitionen abzurufen. Die Gesamtzahl der Segmente, die gleichzeitig ausgeführt werden können, liegt zwischen 1 und 10. Der Standardwert ist 5,

was eine empfohlene Einstellung ist. Sie können dies ändern, indem Sie die Eigenschaft `aws.glue.partition.num.segments` in der `hive-site`-Konfigurationsklassifizierung angeben. Wenn eine Drosselung auftritt, können Sie das Feature ausschalten, indem Sie den Wert auf 1 ändern. Weitere Informationen finden Sie unter [AWS -Glue-Segmentstruktur](#).

Bekannte Probleme

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
- Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine `ConnectSSH`.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
 - Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie unter `http://` eine Verbindung zur Tez-Benutzeroberfläche herstellen `MasterDNS:8080/tez-ui` über eine SSH Verbindung zum primären Clusterknoten wird der Fehler „Adaptervorgang fehlgeschlagen – Timeline-Server () ATS ist außer Reichweite“ angezeigt. Entweder ist er

ausgefallen oder nicht aktiviert“ CORS wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server sendet und localhost nicht den Hostnamen des primären Knotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

5.20.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMRS3Select-Anschluss
emrfs	2.29.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP-Endpoint für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.8	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.8	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.8	HBase Befehlszeilenclient.
hbase-rest-server	1.4.8	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.8	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.214	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.214	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.20.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.20.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.19.1

5.19.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
AWS SDK for Java	1.11.433	1,11.433	1,11.393	1,11.393
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

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
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	0.212	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

5.19.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.19.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.1.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.28.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.8.5-amzn-0	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.5-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.7	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.7	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.7	HBaseBefehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.7	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.7	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.212	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.212	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.19.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml` B. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.19.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN container-log4j.properties von Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>hive-exec-log4j2.properties</code> von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>llap-daemon-log4j2.properties</code> von Hive.
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_conf.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.19.0

5.19.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
AWS SDK for 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

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
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
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.210	0.210	0.206
Spark	2.3.2	2.3.2	2.3.2	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.7.3
ZooKeeper	3.4.13	3.4.12	3.4.12	3.4.12

5.19.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.19.0. Änderungen beziehen sich auf Version 5.18.0.

Erste Version: 7. November 2018

Letzte Aktualisierung: 19. November 2018

Upgrades

- 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

Neue Features

- (19. November 2018) EMR Notebooks ist eine verwaltete Umgebung, die auf Jupyter Notebook basiert. Es unterstützt Spark-Magic-Kernel für Spark PySparkSQL, Spark R und Scala. EMRNotebooks können mit Clustern verwendet werden, die mit Amazon EMR Version 5.18.0 und höher erstellt wurden. Weitere Informationen finden Sie unter [Using EMR Notebooks](#) im Amazon EMR Management Guide.

- Der EMRFS S3-optimierte Committer ist verfügbar, wenn Sie Parquet-Dateien mit Spark und schreiben. EMRFS Dieser Committer verbessert die Schreibleistung. Weitere Informationen finden Sie unter [Verwenden Sie den EMRFS S3-optimierten Committer](#).

Änderungen, Verbesserungen und behobene Probleme

- YARN
 - Die Logik, die den Anwendungsmasterprozess auf die Ausführung auf Core-Knoten beschränkt, wurde geändert. Diese Funktion verwendet jetzt die Funktion und Eigenschaften von YARN Knotenbeschriftungen in den Klassifizierungen `yarn-site` und `capacity-scheduler` der Konfiguration. Weitere Informationen finden Sie unter <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- Standard Amazon Linux AMI für Amazon EMR
 - `ruby18`, `php56` und `gcc48` sind nicht mehr standardmäßig installiert. Diese können auf Wunsch mit `yum` installiert werden.
 - Das `aws-sdk` Ruby Gem ist standardmäßig nicht mehr installiert. Es kann, falls gewünscht, mit `gem install aws-sdk` installiert werden. Bestimmte Komponenten können ebenfalls installiert werden. z. B. `gem install aws-sdk-s3`.

Bekannte Probleme

- EMRNotebooks — Unter bestimmten Umständen kann es vorkommen, dass der Notebook-Editor bei mehreren geöffneten Notebook-Editoren keine Verbindung zum Cluster herstellen kann. Löschen Sie in diesem Fall die Browser-Cookies und öffnen Sie die Notebook-Editoren erneut.
- CloudWatch ContainerPending Metrische und automatische Skalierung — (In 5.20.0 behoben) Amazon gibt EMR möglicherweise einen negativen Wert für `ContainerPending` Wenn `ContainerPending` in einer Auto-Scaling-Regel verwendet wird, verhält sich das Auto Scaling nicht wie erwartet. Vermeiden Sie die Verwendung von `ContainerPending` mit automatischer Skalierung.
- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS

5.19.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.1.0	EMRS3Select-Anschluss

Komponente	Version	Beschreibung
emrfs	2.28.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.8.5-amzn-0	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.5-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.5-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.7	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.7	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.7	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.7	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.7	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.212	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.212	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

5.19.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml` B. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.19.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN container-log4j.properties von Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>hive-exec-log4j2.properties</code> von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_conf.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.18.1

5.18.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
AWS SDK for 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

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
HBase	1.4.7	1.4.7	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
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

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
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

5.18.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.18.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.1.0	EMRS3Select-Anschluss
emrfs	2.27.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.7	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.7	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.7	HBase Befehlszeilenclient.
hbase-rest-server	1.4.7	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.7	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.3-amzn-2	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.210	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.210	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Cli-ent.

5.18.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.18.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.18.0

5.18.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
AWS SDK for 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
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

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
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
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

5.18.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.18.0. Änderungen beziehen sich auf Version 5.17.0.

Veröffentlichungsdatum: 24. Oktober 2018

Upgrades

- Flink 1.6.0

- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Neue Features

- Ab Amazon EMR 5.18.0 können Sie das Amazon EMR Artifact Repository verwenden, um Ihren Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die in bestimmten Amazon-Versionen verfügbar sind. EMR Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Änderungen, Verbesserungen und behobene Probleme

- Hive
 - Unterstützung für S3 Select hinzugefügt. Weitere Informationen finden Sie unter [Verwendung von S3 Select mit Hive zur Verbesserung der Leistung](#).
- Presto
 - Unterstützung für [S3 Select](#) Pushdown hinzugefügt. Weitere Informationen finden Sie unter [S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden](#).
- Spark
 - Die standardmäßige log4j-Konfiguration für Spark wurde dahingehend geändert, dass Container-Logs für Spark-Streaming-Aufträge stündlich angezeigt werden. Dadurch wird verhindert, dass Logs für Spark-Streaming-Aufträge mit langer Laufzeit gelöscht werden.

5.18.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.1.0	EMRS3Select-Anschluss
emrfs	2.27.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.7	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.7	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.7	HBase Befehlszeilenclient.
hbase-rest-server	1.4.7	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.7	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP Reverse-Proxy-Server
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.210	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.210	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Cli-ent.

5.18.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.18.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-protos-hive-site	Ändern Sie die Werte in's .xmlHCatalog. protos-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.17.2

5.17.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
AWS SDK for 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
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

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
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
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

5.17.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.17.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.0.0	EMRS3Select-Anschluss
emrfs	2.26.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.4-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.6	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.6	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.6	HBase Befehlszeilenclient.
hbase-rest-server	1.4.6	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.6	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.206	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.206	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

5.17.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.17.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oracook-site	Ändern Sie die Werte in der Datei oracook-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.17.1

5.17.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
AWS SDK for 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

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
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
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.203	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

5.17.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.17.1. Änderungen beziehen sich auf Version 5.17.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

5.17.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.0.0	EMRS3Select-Anschluss
emrfs	2.26.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSDienst auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP-Endpoint für HDFS Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.6	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.6	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.6	HBase Befehlszeilenclient.
hbase-rest-server	1.4.6	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.6	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.3-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.206	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.206	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Cli-ent.

5.17.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. B. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.17.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.17.0

5.17.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
AWS SDK for 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	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
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

5.17.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.17.0. Änderungen beziehen sich auf Version 5.16.0.

Erste Version: 30. August 2018

Upgrades

- Flink 1.5.2

- HBase 1.4.6
- Presto 0.206

Neue Features

- Unterstützung für TensorFlow hinzugefügt. Weitere Informationen finden Sie unter [TensorFlow](#).

Änderungen, Verbesserungen und behobene Probleme

- JupyterHub
 - Unterstützung für Notebook-Persistenz in Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter [Konfigurieren von Persistenz für Notebooks in Amazon S3](#).
- Spark
 - Unterstützung für [S3 Select](#) hinzugefügt. Weitere Informationen finden Sie unter [S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden](#).
- Die Probleme mit den Cloudwatch-Metriken und der automatischen Skalierungsfunktion in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 wurden behoben.

Bekannte Probleme

- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Wenn Sie ein benutzerdefiniertes Amazon Linux verwenden, das auf einem Amazon Linux AMI mit einem Erstellungsdatum vom 11.08.2018 AMI basiert, kann der Oozie-Server nicht gestartet werden. Wenn Sie Oozie verwenden, erstellen Sie eine benutzerdefinierte Version, die auf einer Amazon AMI Linux-ID mit einem anderen Erstellungsdatum AMI basiert. Sie können den folgenden AWS CLI Befehl verwenden, um eine Liste mit Images IDs für alle HVM Amazon Linux-Versionen AMIs mit einer Version 2018.03 zusammen mit dem Veröffentlichungsdatum zurückzugeben, sodass Sie ein geeignetes Amazon Linux AMI als Basis auswählen können. MyRegion Ersetzen Sie es durch Ihre Regionskennung, z. B. 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
```

5.17.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.0.0	EMRS3Select-Anschluss
emrfs	2.26.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS Dienst auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.4-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.6	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.6	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.6	HBase Befehlszeilenclient.
hbase-rest-server	1.4.6	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.6	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.206	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.206	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

5.17.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.17.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.16.1

5.16.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
AWS SDK for 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

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
HBase	1.4.4	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	2.8.3	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

5.16.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.16.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.25.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.4-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.4-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.8.4-amzn-0	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.4	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.4	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.4	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.4	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.4	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.203	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.203	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

5.16.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.16.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.16.0

5.16.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
AWS SDK for 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

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
HBase	1.4.4	1.4.4	1.4.4	1.4.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.2
Hadoop	2.8.4	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0,4,0	0,4,0	0,4,0
MXNet	1.2.0	1.1.0	1.1.0	1.1.0
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

5.16.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.16.0. Änderungen beziehen sich auf Version 5.15.0.

Erste Version: 19. Juli 2018

Upgrades

- 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
- JDBCRedshift-Treiber 1.2.15.1025

Änderungen, Verbesserungen und behobene Probleme

- HBase
 - [HBASERückportiert -20723](#)
- Presto
 - Konfigurationsänderungen zur Unterstützung der Authentifizierung. LDAP Weitere Informationen finden Sie unter [LDAPAuthentifizierung für Presto bei Amazon verwenden EMR](#).

- Spark
 - [Apache Spark Version 2.3.1, verfügbar ab EMR Amazon-Version 5.16.0, adressiert CVE-2018-8024 und -2018-1334. CVE](#) Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

5.16.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.25.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.4-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.4-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.4	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.4	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.4	HBase Befehlszeilenclient.
hbase-rest-server	1.4.4	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.4	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,1,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.203	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.203	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

5.16.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.16.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.15.1

5.15.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
AWS SDK for 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

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
HBase	1.4.4	1.4.4	1.4.2	1.4.2
HCatalog	2.3.3	2.3.3	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
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

5.15.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.15.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.24.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.8.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.4	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.4	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.4	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.4	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.4	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.3-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

5.15.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.15.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.15.0

5.15.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
AWS SDK for 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

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
HBase	1.4.4	1.4.2	1.4.2	1.4.2
HCatalog	2.3.3	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
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

5.15.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.15.0. Änderungen beziehen sich auf Version 5.14.0.

Erste Version: 21. Juni 2018

Upgrades

- Auf 1.4.4 aktualisiert HBase
- Upgrade auf Hive 2.3.3
- Upgrade auf Hue 4.2.0
- Upgrade auf Oozie 5.0.0
- Upgrade auf ZooKeeper 3.4.12.
- Auf AWS SDK 1.11.333 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Hive
 - [Rückportiert -18069 HIVE](#)
- Hue
 - Hue wurde aktualisiert, um sich korrekt bei Livy zu authentifizieren, wenn Kerberos aktiviert ist. Livy wird jetzt unterstützt, wenn Kerberos mit Amazon verwendet wird. EMR
- JupyterHub
 - Aktualisiert, JupyterHub sodass Amazon standardmäßig LDAP Client-Bibliotheken EMR installiert.
 - Ein Fehler im Skript, das selbstsignierte Zertifikate generiert, wurde behoben.

Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

5.15.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.24.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.4	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.4	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.4	HBase Befehlszeilenclient.
hbase-rest-server	1.4.4	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.4	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.3-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.3-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

5.15.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.15.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei YARN container-log4j.properties von Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>hive-exec-log4j2.properties</code> von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.14.2

5.14.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
AWS SDK for 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

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
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
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

5.14.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.14.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn*-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.23.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.2	HBase Befehlszeilenclient.
hbase-rest-server	1.4.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.

Komponente	Version	Beschreibung
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.14.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.14.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.14.1

5.14.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
AWS SDK for 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

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
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
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

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.14.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.14.1. Änderungen beziehen sich auf Version 5.14.0.

Veröffentlichungsdatum: 17. Oktober 2018

Die Standardeinstellung AMI für Amazon wurde aktualisiert EMR, um potenzielle Sicherheitslücken zu beheben.

5.14.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK

Komponente	Version	Beschreibung
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.23.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.2	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.14.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.14.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.14.0

5.14.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
AWS SDK for Java	1.11.297	1,11.297	1,11.297	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
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
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

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
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

5.14.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.14.0. Änderungen beziehen sich auf Version 5.13.0.

Erste Version: 4. Juni 2018

Upgrades

- Upgrade auf Apache Flink 1.4.2
- Apache wurde auf 1.1.0 MXnet aktualisiert
- Apache Sqoop auf 1.4.7 aktualisiert

Neue Features

- JupyterHub Unterstützung hinzugefügt. Weitere Informationen finden Sie unter [JupyterHub](#).

Änderungen, Verbesserungen und behobene Probleme

- EMRFS
 - Die userAgent Zeichenfolge in Anfragen an Amazon S3 wurde aktualisiert und enthält nun die Benutzer- und Gruppeninformationen des aufrufenden Prinzipals. Dies kann zusammen mit AWS CloudTrail Protokollen für eine umfassendere Nachverfolgung von Anfragen verwendet werden.
- HBase
 - Enthalten ist [HBASE-20447](#), das ein Problem behebt, das zu Cache-Problemen führen könnte, insbesondere bei geteilten Regionen.
- MXnet

- OpenCV-Plattformversion.
- Spark
- Wenn Spark Parquet-Dateien mit an einen Amazon S3 S3-Speicherort schreibtEMRFS, wurde der FileOutputCommitter Algorithmus aktualisiert, sodass er Version 2 statt Version 1 verwendet. Dies reduziert die Anzahl der Umbenennungen, was die Anwendungsleistung verbessert. Diese Änderung beeinträchtigt nicht:
 - Andere Anwendungen als Spark.
 - Anwendungen, die in andere Dateisysteme schreiben, z. B. HDFS (die immer noch Version 1 von verwenden FileOutputCommitter).
 - Anwendungen, die andere Ausgabeformate wie Text oder CSV verwenden, die bereits EMRFS Direct Write verwenden.

Bekannte Probleme

- JupyterHub
 - Die Verwendung von Konfigurationsklassifizierungen zur Einrichtung JupyterHub und individuellen Einrichtung von Jupyter-Notebooks beim Erstellen eines Clusters wird nicht unterstützt. Bearbeiten Sie die Dateien `jupyterhub_config.py` und `jupyter_notebook_config.py` für jeden Benutzer manuell. Weitere Informationen finden Sie unter [Konfiguration JupyterHub](#).
 - JupyterHub kann auf Clustern innerhalb eines privaten Subnetzes nicht gestartet werden und schlägt mit der Meldung fehl. `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` Dies wird durch einen Fehler im Skript verursacht, das selbstsignierte Zertifikate generiert. Verwenden Sie die folgende Problemumgehung zum Generieren selbstsignierter Zertifikate. Alle Befehle werden ausgeführt, während Sie mit dem Primärknoten verbunden sind.

1. Kopieren Sie das Skript zur Zertifikatsgenerierung aus dem Container auf den Primärknoten:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Verwenden Sie einen Texteditor, um Zeile 23 zu ändern, um den öffentlichen Hostnamen in den lokalen Hostnamen zu ändern, wie unten gezeigt:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Führen Sie das Skript aus, um selbstsignierte Zertifikate zu generieren:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Verschieben Sie die Zertifikatsdateien, die das Skript generiert, in das `/etc/jupyter/conf/`-Verzeichnis:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Sie können `tail` die `jupyter.log` Datei überprüfen, um zu überprüfen, ob sie JupyterHub neu gestartet wurde und einen 200-Antwortcode zurückgibt. Beispielsweise:

```
tail -f /var/log/jupyter/jupyter.log
```

Dadurch sollte eine Antwort ähnlich der folgenden ausgegeben:

```
# [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
```

- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

5.14.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.23.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.8.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.2	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.14.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.14.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändern Sie die Werte in der Datei <code>YARN container-log4j.properties</code> von Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von <code>Flink log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.13.1

5.13.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
AWS SDK for Java	1.11.297	1,11.297	1.11.267	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
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
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

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
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

5.13.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.13.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.22.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.8.3-amzn-0	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.2	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.13.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.13.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.13.0

5.13.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
AWS SDK for Java	1.11.297	1.11.267	1.11.267	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.188	0.188	0.188
Spark	2.3.0	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.13.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.13.0. Änderungen beziehen sich auf Version 5.12.0.

Upgrades

- Upgrade auf Spark 2.3.0
- Auf 1.4.2 aktualisiert HBase
- Upgrade auf Presto 0.194
- Auf AWS SDK for Java 1.11.297 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Hive
 - [Rückportiert -15436HIVE](#). Hive APIs wurde dahingehend verbessert, dass nur Ansichten zurückgegeben werden.

Bekannte Probleme

- MXNethat derzeit keine OpenCV-Bibliotheken.

5.13.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.22.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.2	HBase Befehlszeilenclient.
hbase-rest-server	1.4.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.2-amzn-2	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.13.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.13.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase

Klassifizierungen	Beschreibung
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.

Klassifizierungen	Beschreibung
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.12.3

5.12.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
AWS SDK for 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

5.12.3 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.12.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTPEndpunkt für HDFS Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.3-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.0	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.12.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.12.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.12.2

5.12.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
AWS SDK for 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

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
HBase	1.4.0	1.4.0	1.4.0	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
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-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
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

5.12.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.12.2. Änderungen beziehen sich auf Version 5.12.1

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

5.12.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.8.3-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.8.3-amzn-0	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.0	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.0	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.0	HBase-Befehlszeilenclient.
hbase-rest-server	1.4.0	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.12.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.12.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.12.1

5.12.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
AWS SDK for Java	1.11.267	1.11.267	1.11.238	1,11.238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
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

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.12.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.12.1. Änderungen beziehen sich auf Version 5.12.0.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert EMR, um potenzielle Sicherheitslücken zu beheben.

5.12.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.0	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.12.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. B. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.12.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-orahome-site	Ändern Sie die Werte in der Datei orahome-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.12.0

5.12.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
AWS SDK for 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

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
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
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

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
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

5.12.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.12.0. Änderungen beziehen sich auf Version 5.11.1.

Upgrades

- AWS SDK für Java 1.11.238 ⇒ 1.11.267. Weitere Informationen finden Sie im Change Log on [AWS SDK für Java](#). GitHub
- Hadoop 2.7.3 ⇒ 2.8.3. Weitere Informationen finden Sie unter [Apache Hadoop im Releases](#).
- Flink 1.3.2 ⇒ 1.4.0. Weitere Informationen finden Sie in der [Versionsankündigung zu Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Weitere Informationen finden Sie in der [HBase Versionsankündigung](#).
- Hue 4.0.1 ⇒ 4.1.0. Weitere Informationen finden Sie in den [Versionshinweisen](#).
- MxNet 0.12.0 ⇒ 1.0.0. Weitere Informationen finden Sie im [MXNetChange](#) Log on. GitHub
- Presto 0.187 ⇒ 0.188. Weitere Informationen finden Sie in den [Versionshinweisen](#).

Änderungen, Verbesserungen und behobene Probleme

- Hadoop
 - Die `yarn.resourcemanager.decommissioning.timeout`-Eigenschaft wurde zu `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs` geändert. Sie können diese Eigenschaft verwenden, um die Cluster-Skalierung anzupassen. Weitere Informationen finden Sie unter [Cluster Scale-Down](#) im Amazon EMR Management Guide.
 - Hadoop CLI hat die `-d` Option zum Befehl `cp` (copy) hinzugefügt, der das direkte Kopieren spezifiziert. Sie können dies verwenden, um die Erstellung einer `.COPYING`-Zwischendatei zu

vermeiden, wodurch das Kopieren von Daten zwischen Amazon S3 beschleunigt wird. Weitere Informationen finden Sie unter [HADOOP-12384](#).

- Pig
 - Die `pig-env`-Konfigurationsklassifizierung wurde hinzugefügt, die die Konfiguration der Pig-Umgebungseigenschaften vereinfacht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).
- Presto
 - Die Konfigurationsklassifizierung `presto-connector-redshift` wurde hinzugefügt, mit der Sie Werte in der Presto `redshift.properties` konfigurieren können. Weitere Informationen finden Sie unter [Redshift Connector](#) in der Presto-Dokumentation und [Anwendungen konfigurieren](#).
 - Presto-Unterstützung für EMRFS wurde hinzugefügt und ist die Standardkonfiguration. Frühere EMR Amazon-Versionen verwendeten `PrestoS3FileSystem`, was die einzige Option war. Weitere Informationen finden Sie unter [EMRFS und PrestoS3-Konfiguration FileSystem](#).

Note

Wenn Sie die zugrunde liegenden Daten in Amazon S3 mit Amazon EMR Version 5.12.0 abfragen, können Presto-Fehler auftreten. Dies liegt daran, dass Presto die Konfigurationsklassifizierungswerte von `emrfs-site.xml` nicht abrufen kann. Um dieses Problem zu umgehen, erstellen Sie ein `emrfs`-Unterverzeichnis unter `usr/lib/presto/plugin/hive-hadoop2/` und erstellen einen Symlink `usr/lib/presto/plugin/hive-hadoop2/emrfs` zu der vorhandenen Datei `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml`. Starten Sie dann den Presto-Server-Prozess neu (`sudo presto-server stop` gefolgt von `sudo presto-server start`).

- Spark
 - Backported [SPARK-22036](#): Die Multiplikation gibt manchmal Null zurück. `BigDecimal`

Bekanntere Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- SparkR ist nicht für Cluster verfügbar, die mit einem benutzerdefinierten System erstellt wurden. AML, da R nicht standardmäßig auf Clusterknoten installiert ist.

5.12.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.4.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.4.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.4.0	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.4.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0- -1.4 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,13,0- -1,4 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.12.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.12.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.11.4

5.11.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
AWS SDK for 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

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
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
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
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

5.11.4 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.11.4 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-6	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-6	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.11.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.11.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xmlHCatalog.proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Klassifizierungen	Beschreibung
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.11.3

5.11.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
AWS SDK for 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-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
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-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
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

5.11.3 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.11.3. Änderungen beziehen sich auf Version 5.11.2.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

5.11.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-6	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.11.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. B. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.11.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.11.2

5.11.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
AWS SDK for 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

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
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	-	-	-	-
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

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
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

5.11.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.11.2. Änderungen beziehen sich auf Version 5.11.1.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

5.11.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `aufgeführt2.2-amzn-2`.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-6	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.11.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.11.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Klassifizierungen	Beschreibung
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.11.1

5.11.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
AWS SDK for 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

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
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	-	-	-	-
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-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
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

5.11.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.11.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.8.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

5.11.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel

mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-6	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.11.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.11.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.11.0

5.11.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
AWS SDK for 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	Nicht nachverfolgt
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

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
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	2.3.2	2.3.1	2.3.1	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.1	2.3.1	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.184
Spark	2.2.1	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
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

5.11.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.11.0. Änderungen beziehen sich auf Version 5.10.0.

Upgrades

- Hive 2.3.2
- Spark 2.2.1
- SDK für Java 1.11.238

Neue Features

- Spark
 - Die Einstellung `spark.decommissioning.timeout.threshold` wurde hinzugefügt, was das Verhalten von Spark bei der Außerbetriebnahme bei Verwendung von Spot Instances verbessert. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
 - Die `aws-sagemaker-spark-sdk` Komponente wurde zu Spark hinzugefügt, die Amazon SageMaker Spark und zugehörige Abhängigkeiten für die Spark-Integration mit [Amazon](#) installiert SageMaker. Sie können Amazon SageMaker Spark verwenden, um Spark-Pipelines für maschinelles Lernen (ML) mithilfe von SageMaker Amazon-Stufen zu erstellen. Weitere Informationen finden Sie in der [SageMaker Spark-Readme-Datei](#) zu Apache Spark GitHub und zur [Verwendung von Apache Spark mit Amazon SageMaker](#) im Amazon SageMaker Developer Guide.

Bekannte Probleme

- MXNetenthält keine OpenCV-Bibliotheken.
- Hive 2.3.2 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

5.11.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark SDK

Komponente	Version	Beschreibung
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-6	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-6	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.11.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.11.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraooop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Klassifizierungen	Beschreibung
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.10.1

5.10.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
AWS SDK for Java	1.11.221	1.11.221	1.11.183	1.11,183

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Python	2,7, 3,4	2,7, 3,4	Nicht nachverfolgt	Nicht nachverfolgt
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	2.3.1	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

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
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
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

5.10.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.10.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich

EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `amzn-2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-5	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.1-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.1-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.10.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.10.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat

Klassifizierungen	Beschreibung
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.10.0

5.10.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
AWS SDK for Java	1.11.221	1.11.183	1.11,183	1.11.160
Python	2,7, 3,4	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
MXNet	0.12.0	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.16.0
Presto	0.187	0.184	0.184	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.10.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.10.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.9.0.

Upgrades

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

Neue Features

- Die Kerberos-Authentifizierung wird jetzt unterstützt. Weitere Informationen finden Sie unter [Verwenden der Kerberos-Authentifizierung](#) im Amazon EMR Management Guide.
- Unterstützung für IAM Rollen für hinzugefügt. EMRFS Weitere Informationen finden [Sie unter IAM Rollen für EMRFS Anfragen an Amazon S3 konfigurieren](#) im Amazon EMR Management Guide
- Unterstützung für GPU basierte P2- und P3-Instance-Typen hinzugefügt. Weitere Informationen finden Sie unter [Amazon EC2 P2-Instances](#) und [Amazon EC2 P3-Instances](#). NVIDIA Die Treiber 384.81 und der CUDA Treiber 9.0.176 sind standardmäßig auf diesen Instance-Typen installiert.
- Unterstützung für [Apache MXNet](#) hinzugefügt.

Änderungen, Verbesserungen und behobene Probleme

- Presto
 - Unterstützung für die Verwendung des AWS Glue-Datenkatalogs als Standard-Hive-Metastore hinzugefügt. Weitere Informationen finden Sie unter [Presto mit dem AWS Glue-Datenkatalog verwenden](#).
 - Unterstützung von [koordinatenbasierten Funktionen](#) hinzugefügt.
 - Unterstützung eines [Überlaufs auf die Festplatte](#) für Joins hinzugefügt.
 - Unterstützung für den [Redshift-Connector](#) hinzugefügt.
- Spark
 - [SPARK-20640](#) wurde zurückportiert, wodurch das RPC-Timeout und die Wiederholungsversuche für Shuffle-Registrierungswerte mithilfe von Eigenschaften konfiguriert werden können. `spark.shuffle.registration.timeout`
`spark.shuffle.registration.maxAttempts`
 - [SPARK-21549](#) wurde zurückportiert, wodurch ein Fehler behoben wird, der beim Schreiben von benutzerdefinierten Daten an andere Speicherorte auftritt. OutputFormat HDFS
- Rückportierung von [Hadoop-13270](#)
- Die Bibliotheken Numpy, Scipy und Matplotlib wurden aus dem Basis-Amazon entfernt. EMR AMI Wenn diese Bibliotheken für Ihre Anwendung erforderlich sind, werden sie im folgenden Repository bereitgestellt, sodass Sie sie unter Verwendung einer Bootstrap-Aktion mit `yum install` auf allen Knoten installieren können.

- In der EMR Amazon-Basis sind AMI keine RPM Anwendungspakete mehr enthalten, sodass die RPM Pakete nicht mehr auf Clusterknoten vorhanden sind. Custom AMIs und Amazon EMR Base verweisen AMI jetzt auf das RPM Paket-Repository in Amazon S3.
- Aufgrund der Einführung der sekundengenauen Abrechnung in Amazon EC2 lautet das Standardverhalten beim Herunterfahren jetzt bei Abschluss der Aufgabe beenden und nicht mehr bei Instance-Stunde beenden. Weitere Informationen finden Sie unter [Konfigurieren eines Cluster-Scale-Downs](#).

Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.1 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

5.10.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-5	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-5	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-5	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.1-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.1-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.10.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.10.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Klassifizierungen	Beschreibung
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.9.1

5.9.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die Anwendungsversionen aufgeführt, die in dieser Version von Amazon verfügbar sind, EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend).

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
AWS SDK for Java	1.11.183	1.11,183	1.11.160	1,11.160

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	3.12.0	3.12.0
Iceberg	-	-	-	-
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

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
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

5.9.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.9.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich

EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.19.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-4	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBaseBefehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	RESTSchnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mysql-server	5.5.54+	Mein SQL Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.184	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.184	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.9.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.9.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in <code>.xmlHCatalog.proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Klassifizierungen	Beschreibung
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.9.0

5.9.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
AWS SDK for Java	1.11.183	1.11.160	1,11.160	1,11.160

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-
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

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
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

5.9.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.9.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.8.0.

Veröffentlichungsdatum: 5. Oktober 2017

Letzte Aktualisierung der Funktionen: 12. Oktober 2017

Upgrades

- AWS SDK for Java Version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Neue Features

- Livy-Support hinzugefügt (Version 0.4.0-incubating). Weitere Informationen finden Sie unter [Apache Livy](#).
- Support für Hue-Notebooks für Spark hinzugefügt.
- Unterstützung für EC2 Amazon-Instances der i3-Serie hinzugefügt (12. Oktober 2017).

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Neue Funktionen hinzugefügt, die sicherstellen, dass Spark die Knotenbeendigung infolge einer manuellen Größenänderung oder einer Anforderung einer Auto Scaling-Richtlinie zuverlässiger ausführt. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
 - SSL wird anstelle von 3 DES für die Verschlüsselung während der Übertragung für den Blocktransfer-Service verwendet, wodurch die Leistung bei der Verwendung von EC2 Amazon-Instance-Typen mit AES -NI verbessert wird.
 - [Rückportiert -21494SPARK](#).
- Zeppelin
 - [Rückportiert -2377. ZEPPELIN](#)
- HBase
 - Patch [HBASE-18533](#) hinzugefügt, der zusätzliche Werte für die Konfiguration mithilfe der HBase BucketCache Konfigurationsklassifizierung ermöglicht. `hbase-site`
- Hue
 - AWS Glue Data Catalog-Unterstützung für den Hive-Abfrageeditor in Hue hinzugefügt.
 - Standardmäßig können Superuser in Hue auf alle Dateien zugreifen, auf die EMR IAM Amazon-Rollen zugreifen dürfen. Neu erstellte Benutzer verfügen nicht automatisch über Zugriffsberechtigungen auf den Amazon-S3-Filebrowser und für ihre Gruppen müssen die `filebrowser.s3_access`-Berechtigungen aktiviert sein.
 - Es wurde ein Problem behoben, das dazu führte, dass auf die zugrunde liegenden JSON Daten, die mit dem AWS Glue-Datenkatalog erstellt wurden, nicht zugegriffen werden konnte.

Bekannte Probleme

- Der Clusterstart schlägt fehl, wenn alle Anwendungen installiert sind und die Standardgröße des EBS Amazon-Root-Volumes nicht geändert wird. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren `--ebs-root-volume-size` Parameter an.
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle `LOCATION` hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

5.9.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.19.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-4	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-4	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein SQL Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.184	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.184	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.9.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.9.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-prot-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oracook-site	Ändern Sie die Werte in der Datei oracook-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.8.3

5.8.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
AWS SDK for Java	1.11.160	1,11.160	1,11.160	1,11.160
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
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	2.3.0
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
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	-	-	-	-

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
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

5.8.3 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.8.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-3	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.8.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.8.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-prot-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.8.2

5.8.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
AWS SDK for Java	1.11.160	1,11.160	1,11.160	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.1.1

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
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
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

5.8.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.8.2. Änderungen beziehen sich auf Version 5.8.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert EMR, um potenzielle Sicherheitslücken zu beheben.

5.8.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-3	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.8.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.8.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.8.1

5.8.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
AWS SDK for Java	1.11.160	1,11.160	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.0	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.1.1	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

5.8.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.8.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.8.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

5.8.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-3	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.8.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.8.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.8.0

5.8.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
AWS SDK for Java	1.11.160	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
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	4.9.0
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

5.8.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.8.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.7.0.

Erste Version: 10. August 2017

Letzte Aktualisierung der Funktionen: 25. September 2017

Upgrades

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache Hive-Website.
- Spark 2.2.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache Spark-Website.

Neue Features

- Support zur Anzeige des Anwendungsverlaufs hinzugefügt (25. September 2017). Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Integration mit dem AWS Glue-Datenkatalog
 - Hive und Spark haben jetzt die Möglichkeit SQL, AWS Glue Data Catalog als Hive-Metadatenpeicher zu verwenden. Weitere Informationen finden Sie unter [Den AWS Glue-Datenkatalog als Metastore für Hive verwenden](#) und [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).
- Der Anwendungsverlauf wurde zu den Cluster-Details hinzugefügt, sodass Sie historische Daten für YARN Anwendungen und zusätzliche Details für Spark-Anwendungen anzeigen können. Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Amazon EMR Management Guide.
- Oozie
 - Rückportiert [OOZIE-2748](#).
- Hue
 - Rückportiert [HUE-5859](#)
- HBase
 - Es wurde ein Patch hinzugefügt, um die Startzeit des HBase Masterservers mithilfe von Java Management Extensions () JMX anzuzeigen. `getMasterInitializedTime`

- Patch zur Verbesserung der Cluster-Startzeit hinzugefügt.

Bekannte Probleme

- Der Clusterstart schlägt fehl, wenn alle Anwendungen installiert sind und die Standardgröße des EBS Amazon-Root-Volumes nicht geändert wird. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren `--ebs-root-volume-size` Parameter an.
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

- Spark – Bei der Verwendung von Spark tritt bei dem apppusher-Daemon bei Spark-Aufträgen mit langer Ausführungsdauer nach einigen Stunden oder Tagen möglicherweise ein Speicherleck-Problem bei dem Datei-Handler auf. Um das Problem zu beheben, stellen Sie eine Verbindung mit dem Primärknoten her und geben `sudo /etc/init.d/appusher stop` ein. Dadurch wird der Appusher-Daemon gestoppt, den Amazon EMR automatisch neu startet.
- Application history
 - Es sind keine Verlaufsdaten für beendete Spark-Executors verfügbar.
 - Der Anwendungsverlauf ist nicht für Cluster verfügbar, die eine Sicherheitskonfiguration verwenden, um die Verschlüsselung bei der Übertragung zu aktivieren.

5.8.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.8.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.8.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xmlHCatalog.proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.7.1

5.7.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.3.0	1.2.1	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
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

5.7.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.7.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP-Endpoint für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltH Catalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungseengine für. YARN

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTPApache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.7.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.7.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.7.0

5.7.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.2.1	1.2.1	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
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

5.7.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.7.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.6.0.

Veröffentlichungsdatum: 13. Juli 2017

Upgrades

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Neue Features

- Es wurde die Möglichkeit hinzugefügt, ein customAmazon Linux anzugebenAMI, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Benutzerdefiniert verwenden AMI](#).

Änderungen, Verbesserungen und behobene Probleme

- HBase
 - Es wurde die Möglichkeit hinzugefügt, HBase Read-Replica-Cluster zu konfigurieren. Weitere Informationen finden Sie unter [Verwenden eines Read Replica-Clusters](#)
 - Mehrere Fehlerbehebungen und Verbesserungen
- Presto – neue Konfigurationsfunktion `node.properties` hinzugefügt.
- YARN- Fähigkeit zur Konfiguration hinzugefügt `container-log4j.properties`
- Sqoop - rückportiert [SQOOP-2880](#), was ein Argument einführt, mit dem Sie das temporäre Sqoop-Verzeichnis festlegen können.

5.7.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0- -1.3 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0- -1,3 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.7.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.7.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.6.1

5.6.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.6.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.6.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.1	HBase Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.6.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.6.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.6.0

5.6.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

5.6.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.6.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.5.0.

Veröffentlichungsdatum: 5. Juni 2017

Upgrades

- Flink 1.2.1
- HBase1.3.1
- Mahout 0.13.0. Dies ist die erste Version von Mahout, die Spark 2.x in EMR Amazon-Version 5.0 und höher unterstützt.
- Spark 2.1.1

Änderungen, Verbesserungen und behobene Probleme

- Presto
 - Es wurde die Möglichkeit hinzugefügt, die Kommunikation zwischen Presto-Knoten zu SSL aktivieren/zu TLS sichern, indem die Verschlüsselung während der Übertragung mithilfe einer Sicherheitskonfiguration aktiviert wird. Weitere Informationen finden Sie unter [Verschlüsselung von Daten während der Übertragung](#).
 - [Presto 7661](#) wurde rückportiert, wodurch der Anweisung EXPLAIN ANALYZE die Option VERBOSE hinzugefügt wurde, um detailliertere Low-Level-Statistiken über einen Abfrageplan bereitzustellen.

5.6.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-2	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.1	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.6.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.6.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-protos-hive-site	Ändern Sie die Werte in's .xmlHCatalog. protos-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.5.4

5.5.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
Amazon CloudWatch Agent	-	-	-	-
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

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
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
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

5.5.4 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.5.4 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.0	HBase Befehlszeilenclient.
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.5.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.5.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.5.3

5.5.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
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
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

5.5.3 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.5.3. Änderungen beziehen sich auf Version 5.5.2.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

5.5.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.3.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.0	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.5.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.5.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat

Klassifizierungen	Beschreibung
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.5.2

5.5.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
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	3.11.0
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

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
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

5.5.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.5.2. Änderungen beziehen sich auf Version 5.5.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert EMR, um potenzielle Sicherheitslücken zu beheben.

5.5.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.1.1-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Client.

5.5.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.5.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase

Klassifizierungen	Beschreibung
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.

Klassifizierungen	Beschreibung
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.5.1

5.5.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
Amazon CloudWatch Agent	-	-	-	-
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

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
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
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

5.5.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.5.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.5.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

5.5.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.0	HBase Befehlszeilenclient.
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.5.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.5.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.5.0

5.5.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
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

5.5.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.5.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.4.0.

Veröffentlichungsdatum: 26. April 2017

Upgrades

- Hue 3.12

- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Rückportierter Spark-Patch ([SPARK-20115](#)) Es wurde behoben, DAGScheduler um alle verlorenen Shuffle-Blöcke neu zu berechnen, wenn der externe Shuffle-Dienst für Version 2.1.0 von Spark, die in dieser Version enthalten ist, nicht verfügbar ist.
- Flink
 - Flink ist nun in Scala 2.11 integriert. Wenn Sie Scala API und Bibliotheken verwenden, empfehlen wir Ihnen, Scala 2.11 in Ihren Projekten zu verwenden.
 - Ein Problem wurde behoben, bei dem die Standardeinstellungen HADOOP_CONF_DIR und YARN_CONF_DIR nicht ordnungsgemäß festgelegt wurden, sodass `start-scala-shell.sh` nicht funktionierte. Außerdem wurde die Möglichkeit hinzugefügt, diese Werte mit `env.hadoop.conf.dir` und `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` oder der Konfigurationsklassifizierung `flink-conf` festzulegen.
 - Es wurde ein neuer EMR -spezifischer Befehl `flink-scala-shell` als Wrapper für `start-scala-shell.sh` eingeführt. Wir empfehlen, diesen Befehl statt `start-scala-shell` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-scala-shell -n 2` eine Flink Scala-Shell mit einer Aufgabenparallelität von 2.
 - Es wurde ein neuer EMR -spezifischer Befehl `flink-yarn-session` als Wrapper für `yarn-session.sh` eingeführt. Wir empfehlen, diesen Befehl statt `yarn-session` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-yarn-session -d -n 2` eine langlebige Flink-Sitzung in einem getrennten Status mit zwei Aufgabenmanagern.
 - Der adressierte ([FLINK-6125](#)) Commons-HttpClient ist in Flink 1.2 nicht mehr schattiert.
- Presto
 - Unterstützung für Authentifizierung hinzugefügt. LDAP Für die Verwendung LDAP mit Presto bei Amazon EMR müssen Sie den HTTPS Zugriff für den Presto-Koordinator (`http-server.https.enabled=true` in `config.properties`) aktivieren. Einzelheiten zur Konfiguration finden Sie in der [LDAPPresto-Dokumentation unter Authentifizierung](#).
 - Unterstützung für `SHOW GRANTS` hinzugefügt.

- Amazon EMR Base Linux AMI
 - EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie in den [Versionshinweisen zu Amazon Linux AMI 2017.03](#).
 - Python 2.6 wurde aus dem EMR Linux-Basisimage von Amazon entfernt. Python 2.7 und 3.4 sind standardmäßig installiert. Sie können Python 2.6 bei Bedarf manuell installieren.

5.5.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.0	HBase Befehlszeilenclient.
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilen-Cli-ent.

5.5.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.5.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.4.1

5.4.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.166	0.166	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
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

5.4.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.4.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.15.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.3.0	HBase Befehlszeilenclient.
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.166	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.166	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

5.4.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.4.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.4.0

5.4.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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
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	-	-	-	-

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
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
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

5.4.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.4.0. Die Änderungen beziehen sich auf die Version EMR 5.3.0 von Amazon.

Veröffentlichungsdatum: 8. März 2017

Upgrades

- Upgrade auf Flink 1.2.0

- Auf 1.3.0 aktualisiert HBase
- Upgrade auf Phoenix 4.9.0

Note

Wenn Sie ein Upgrade von einer früheren Version von Amazon EMR auf EMR Amazon-Version 5.4.0 oder höher durchführen und die sekundäre Indizierung verwenden, aktualisieren Sie lokale Indizes wie in der [Apache](#) Phoenix-Dokumentation beschrieben. Amazon EMR entfernt die erforderlichen Konfigurationen aus der `hbase-site` Klassifizierung, aber die Indizes müssen neu gefüllt werden. Online- und Offline-Upgrades von Indizes werden unterstützt. Online-Upgrades sind die Standardeinstellung. Dies bedeutet, dass Indizes neu gefüllt werden, während die Initialisierung von Phoenix-Clients Version 4.8.0 oder höher ausgeführt wird. Um Offline-Upgrades anzugeben, setzen Sie die `phoenix.client.localIndexUpgrade` Konfiguration in der `phoenix-site` Klassifizierung auf `False` und dann SSH auf den primären Knoten, der ausgeführt werden soll. `psql [zookeeper] -1`

- Upgrade auf Presto 0.166
- Upgrade auf Zeppelin 0.7.0

Änderungen und Verbesserungen

- Es werden jetzt r4-Instances unterstützt. Siehe [EC2Amazon-Instance-Typen](#).

5.4.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich. EMR werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.15.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpoint für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.3.0	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilenclient.
hbase-rest-server	1.3.0	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.3.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,9,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.166	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.166	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.7.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

5.4.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.4.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-prot-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.3.2

5.3.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.4	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.10.0
Iceberg	-	-	-	-

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
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
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

5.3.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.3.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.4	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt H Catalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

5.3.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.3.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.3.1

5.3.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
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

5.3.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.3.1. Die Änderungen beziehen sich auf die Version EMR 5.3.0 von Amazon.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen am Backport von Zeppelin-Patches und an der Aktualisierung der Standardeinstellung AMI für Amazon. EMR

5.3.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.1.4	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Client.

5.3.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.3.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändern HBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xml</code> HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der <code>spark-defaults.conf</code> -Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.3.0

5.3.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
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

5.3.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.3.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.2.1.

Veröffentlichungsdatum: 26. Januar 2017

Upgrades

- Upgrade auf Hive 2.1.1
- Upgrade auf Hue 3.11.0
- Upgrade auf Spark 2.1.0
- Upgrade auf Oozie 4.3.0
- Upgrade auf Flink 1.1.4

Änderungen und Verbesserungen

- Ein Patch für Hue wurde hinzugefügt, mit dem Sie die Einstellung `interpreters_shown_on_wheel` verwenden können, um zu konfigurieren, was Interpreter in der Notebook-Auswahl als Erstes anzeigen, unabhängig von ihrer Reihenfolge in der Datei `hue.ini`.
- Die Konfigurationsklassifizierung `hive-parquet-logging` wurde hinzugefügt, mit der Sie Werte in der Hive-Datei `parquet-logging.properties` konfigurieren können.

5.3.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.1.4	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.1-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.1.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Client.

5.3.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.3.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.

Klassifizierungen	Beschreibung
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xmlHCatalog.proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat <code>log4j2.properties</code> von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei <code>webhcat-site.xml</code> von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der <code>beeline-log4j2.properties</code> -Datei in Hive.
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der <code>spark-defaults.conf</code> -Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.2.3

5.2.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.8

5.2.3 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.2.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

5.2.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.2.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.2.2

5.2.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
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

5.2.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.2.2. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.2.1.

Veröffentlichungsdatum: 2. Mai 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Backported [SPARK-194459](#), wodurch ein Problem behoben wurde, bei dem das Lesen aus einer ORC Tabelle mit char/varchar-Spalten fehlschlagen kann.

5.2.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httfs-server	2.7.3-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Client.

5.2.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.2.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.

Klassifizierungen	Beschreibung
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.2.1

5.2.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
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
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

5.2.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.2.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.2.0.

Veröffentlichungsdatum: 29. Dezember 2016

Upgrades

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf ZooKeeper 3.4.9. Weitere Informationen finden Sie in den [ZooKeeperVersionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

Änderungen und Verbesserungen

- Unterstützung für den Instance-Typ Amazon EC2 m4.16xlarge in EMR Amazon-Version 4.8.3 und höher hinzugefügt, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- Der Speicherort von Flink und die YARN Konfigurationspfade sind jetzt standardmäßig festgelegt/etc/default/flink, sodass Sie die Umgebungsvariablen nicht festlegen müssen FLINK_CONF_DIR und HADOOP_CONF_DIR wenn Sie die `flink` oder `yarn-session.sh` - Treiberskripte ausführen, um Flink-Jobs zu starten.
- Unterstützung für FlinkKinesisConsumer Klassen hinzugefügt.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem `ControlledJob #` mit einer Nullzeiger-Ausnahme (NPE) `toString` fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

5.2.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.1.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

5.2.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.2.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-prot-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.2.0

5.2.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
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
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

5.2.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.2.0. Die Änderungen beziehen sich auf die Version EMR 5.1.0 von Amazon.

Veröffentlichungsdatum: 21. November 2016

Änderungen und Verbesserungen

- Amazon S3 S3-Speichermodus für hinzugefügtHBase.
- Ermöglicht es Ihnen, einen Amazon S3 S3-Speicherort für das HBase Rootdir anzugeben. Weitere Informationen finden Sie [HBaseunter Amazon S3](#).

Upgrades

- Upgrade auf Spark 2.0.2

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem /mnt bei Instance-Typen EBS nur auf 2 TB beschränkt war.
- Es wurde ein Problem mit Instance-Controller und Logpusher-Protokollen behoben, die an die entsprechenden OUT-Dateien anstelle ihrer üblichen log4j-konfigurierten LOG-Dateien, die stündlich rotieren, ausgegeben wurden. Da die OUT-Dateien nicht rotieren, ist letztendlich die /emr-Partition belegt. Dieses Problem betrifft nur Hardware-Instanztypen für virtuelle Maschinen (HVM).

5.2.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.12.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-0	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-0	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

5.2.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.2.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Amazon EMR — kuratierte Einstellungen für Apache. HBase
hbase-env	Werte in der Umgebung ändernHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-prot-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.

Klassifizierungen	Beschreibung
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.1.1

5.1.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
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
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

5.1.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

⚠ Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

5.1.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.11.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-0	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-0	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

5.1.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.1.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändern Sie die Werte in der UmgebungHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.

Klassifizierungen	Beschreibung
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.1.0

5.1.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.150	0.150
Spark	2.0.1	2.0.1	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
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

5.1.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.1.0. Die Änderungen beziehen sich auf die Version EMR 5.0.3 von Amazon.

Veröffentlichungsdatum: 03. November 2016

Änderungen und Verbesserungen

- Flink 1.1.3 wird jetzt unterstützt.
- Presto wurde als Option im Notebookabschnitt von Hue hinzugefügt.

Upgrades

- Auf 1.2.3 aktualisiert HBase
- Upgrade auf Zeppelin 0.6.2

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem mit Tez-Abfragen auf Amazon S3 behoben, bei ORC denen Dateien nicht so gut funktionierten wie frühere Amazon EMR 4.x-Versionen.

5.1.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn*-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.11.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.3	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.3	HBase Befehlszeilenclient.
hbase-rest-server	1.2.3	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.3	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

5.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändern Sie die Werte in der UmgebungHBase.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.

Klassifizierungen	Beschreibung
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.0.3

5.0.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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

5.0.3 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.0.3. Die Änderungen beziehen sich auf die Version EMR 5.0.0 von Amazon.

Veröffentlichungsdatum: 24. Oktober 2016

Upgrades

- Upgrade auf Hadoop 2.7.3
- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889

zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Web Interface](#) in der Presto-Dokumentation.

- Upgrade auf Spark 2.0.1
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

5.0.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.10.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS-Befehlszeilenclient und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httfs-server	2.7.3-amzn-0	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
spark-client	2.0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.0.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTPApache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Client.

5.0.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.0.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .

Klassifizierungen	Beschreibung
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.0.2

5.0.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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
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

5.0.2 Versionshinweise

5.0.2 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.2-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.150	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.150	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.0.0	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	2.0.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Cliant.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.1- SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cliant.

5.0.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.0.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.

Klassifizierungen	Beschreibung
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.0.1

5.0.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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

5.0.1 Versionshinweise

5.0.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.1.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.150	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.150	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.0	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	2.0.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.1- SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli ent.

5.0.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.0.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.

Klassifizierungen	Beschreibung
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 5.0.0

5.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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

5.0.0 Versionshinweise

Veröffentlichungsdatum: 27. Juli 2016

Upgrades

- Upgrade auf Hive 2.1
- Upgrade auf Presto 0.150
- Upgrade auf Spark 2.0
- Upgrade auf Hue 3.10.0
- Upgrade auf Pig 0.16.0

- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin 0.6.1

Änderungen und Verbesserungen

- Amazon EMR unterstützt die neuesten Open-Source-Versionen von Hive (Version 2.1) und Pig (Version 0.16.0). Wenn Sie Hive oder Pig EMR in der Vergangenheit bei Amazon verwendet haben, kann sich dies auf einige Anwendungsfälle auswirken. Weitere Informationen finden Sie unter [Hive](#) und [Pig](#).
- Die Standard-Ausführungs-Engine für Hive und Pig ist jetzt Tez. Um dies zu ändern, bearbeiten Sie die entsprechenden Werte in den Konfigurationsklassifizierungen `hive-site` bzw. `pig-properties`.
- Eine verbesserte Schritt-Debugging-Funktion wurde hinzugefügt. Mit dieser Funktion können Sie die Ursache von Schrittfehlern sehen, wenn der Service die Ursache ermitteln kann. Weitere Informationen finden Sie unter [Enhanced Step Debugging](#) im Amazon EMR Management Guide.
- Anwendungen, die zuvor mit "-Sandbox" endeten, tragen diesen Suffix nicht mehr. Dies kann Ihre Automatisierung, zum Beispiel, wenn Sie Skripts zum Starten von Clustern mit diesen Anwendungen verwenden, unterbrechen. Die folgende Tabelle zeigt Anwendungsnamen in Amazon EMR 4.7.2 im Vergleich zu Amazon EMR 5.0.0.

Änderungen der Anwendungsnamen

Amazon EMR 4,7,2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Sandkasten	ZooKeeper

- Spark ist jetzt für Scala 2.11 kompiliert.
- Java 8 ist jetzt der Standard. JVM Alle Anwendungen werden mit der Java 8-Laufzeit ausgeführt. Es sind keine Änderungen am Byte-Codeziel der Anwendungen vorgenommen worden. Die meisten Anwendungen zielen weiterhin auf Java 7 ab.

- Zeppelin enthält jetzt Authentifizierungsfunktionen. Weitere Informationen finden Sie unter [Zeppelin](#).
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

5.0.0 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Befehlszeilenclient und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httfs-server	2.7.2-amzn-3	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.

Komponente	Version	Beschreibung
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.150	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.150	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
spark-client	2.0.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.0	In-Memory-Ausführungsengine für YARN
spark-yarn-slave	2.0.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Clients.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.1- SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Clients.

5.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie `hive-site.xml` z. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-5.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .

Klassifizierungen	Beschreibung
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xml HCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j2	Ändern Sie die Werte in den ebHCat log4j2.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei webhcat-site.xml von HCatalog W. ebHCat
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei hive-exec-log4j2.properties von Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändern Sie die Werte in der Umgebung. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

Amazon EMR 4.x-Release-Versionen

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon EMR 4.x-Release-Version verfügbar sind.

Wenn Sie einen Cluster starten, können Sie aus mehreren Versionen von Amazon wählen EMR. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.2.0`.

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen

Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon EMR 4.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#).

Themen

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)
- [Unterschiede in Amazon-EMR-4.x-Versionen](#)
- [EMRAmazon-Version 4.9.6](#)
- [EMRAmazon-Version 4.9.5](#)
- [EMRAmazon-Version 4.9.4](#)
- [EMRAmazon-Version 4.9.3](#)
- [EMRAmazon-Version 4.9.2](#)
- [EMRAmazon-Version 4.9.1](#)
- [EMRAmazon-Version 4.8.5](#)
- [EMRAmazon-Version 4.8.4](#)
- [EMRAmazon-Version 4.8.3](#)
- [EMRAmazon-Version 4.8.2](#)
- [EMRAmazon-Version 4.8.1](#)
- [EMRAmazon-Version 4.8.0](#)
- [EMRAmazon-Version 4.7.4](#)
- [EMRAmazon-Version 4.7.3](#)
- [EMRAmazon-Version 4.7.2](#)
- [EMRAmazon-Version 4.7.1](#)
- [EMRAmazon-Version 4.7.0](#)
- [EMRAmazon-Version 4.6.1](#)
- [EMRAmazon-Version 4.6.0](#)
- [EMRAmazon-Version 4.5.0](#)
- [EMRAmazon-Version 4.4.0](#)

- [EMRAmazon-Version 4.3.0](#)
- [EMRAmazon-Version 4.2.0](#)
- [EMRAmazon-Version 4.1.0](#)
- [EMRAmazon-Version 4.0.0](#)

Anwendungsversionen in Amazon-EMR-4.x-Versionen

Eine umfassende Tabelle mit den in jeder Amazon-EMR-4.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#) in Ihrem Browser.

Unterschiede in Amazon-EMR-4.x-Versionen

Die Dokumentation für Amazon-EMR-Feature in Amazon EMR gibt die Verwaltungshandbuch-Version an, in der ein Feature verfügbar wurde, sowie zutreffende Unterschiede zwischen Amazon-EMR-Features, die auf Version 4.0.0 zurückreichen.

Beginnend mit Amazon-EMR-Version 5.0.0 erhielten einige Anwendungen ein erhebliches Versionsupgrade, bei dem die Installation oder betriebliche Details geändert wurden. Andere wurden von Sandbox-Anwendungen zu nativen Anwendungen hochgestuft. Jedes Thema in diesem Abschnitt bietet bedeutende anwendungsspezifische Unterschiede bei der Verwendung von Amazon-EMR-4.x-Versionen.

Themen

- [Sandbox-Anwendungen](#)
- [Überlegungen für die Verwendung von Hive in Amazon EMR 4.x](#)
- [Überlegungen für die Verwendung von Pig in Amazon EMR 4.x](#)

Sandbox-Anwendungen

Bei der Verwendung von Amazon-EMR-4.x-Versionen werden einige Anwendungen als Sandbox-Anwendungen betrachtet. Sandbox-Anwendungen sind frühe Versionen der Anwendung, die wir aufgrund der Nachfrage zum Zeitpunkt der ersten Amazon-EMR-Version zur Verfügung gestellt haben. Sie können die Konsole, AWS CLI oder API installieren, damit Amazon-EMR-Sandbox-Anwendungen wie native Anwendungen installiert. Sandbox-Anwendungen erhalten jedoch

eingeschränkten Support und eingeschränkte Dokumentation. Sandbox-Anwendungen wurden in Amazon-EMR-Version 5.0.0 und höher zu nativen, voll unterstützten Anwendungen. Die folgenden Anwendungen sind Sandbox-Anwendungen in Amazon-EMR-4.x-Versionen:

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

Wenn Sie Sandbox-Anwendungen installieren, werden die Anwendungsnamen mit dem Suffix `-sandbox` angegeben. Beispiel: Zum Installieren der Sandbox-Version von *Presto* verwenden Sie `Presto-sandbox`. Die Installation kann länger als bei einer voll unterstützten Anwendung dauern. Die für jede Anwendung in diesem Abschnitt aufgelisteten Versionsnummern entsprechen der Community-Version der Anwendung.

Oozie (Sandbox-Versionen)

Oozie ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.1.0 verfügbar.

Oozie-Beispiele werden von den Sandbox-Versionen nicht standardmäßig installiert. Zum Installieren der Beispiele nutzen Sie SSH für den Master-Knoten und führen Sie `install-oozie-examples` aus.

Informationen zur Oozie-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.9.6	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
		cemanager, oozie-client, oozie-server
emr-4.9.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server
emr-4.9.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.9.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.8.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.8.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.7.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.7.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
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 (Sandbox-Versionen)

Presto ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.1.0 verfügbar.

Informationen zur Presto-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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,

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
		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

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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-resource-manager, 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-resource-manager, 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-resource-manager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
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-resource-manager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Sqoop (Sandbox-Versionen)

Sqoop ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.4.0 verfügbar.

Informationen zur Sqoop-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
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-resource-manager, 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-

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
		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

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
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-httpfs-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
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 (Sandbox-Versionen)

Zeppelin ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.1.0 verfügbar.

Informationen zur Zeppelin-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
		nager, hadoop-yarn-resour cemanager, 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- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, 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- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.9.3	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.2	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.9.1	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.5	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.8.4	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.3	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.8.2	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.1	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.8.0	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.4	0.5.6	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.7.3	0.5.6	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.2	0.5.6	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
emr-4.6.0	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
emr-4.5.0	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
emr-4.4.0	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

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
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 (Sandbox-Versionen)

Zookeeper ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.6.0 verfügbar.

Informationen zur Version von Zookeeper-Sandbox

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
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-resourcemanager, 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-resourcemanager, 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-resourcemanager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.9.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.8.2	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-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.7.1	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-4.7.0	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
emr-4.6.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
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-resource manager, zookeeper-client, zookeeper-server

Überlegungen für die Verwendung von Hive in Amazon EMR 4.x

In diesem Abschnitt werden die Unterschiede behandelt, die Sie berücksichtigen sollten, wenn Sie Hive-Version 1.0.0 auf Amazon-EMR-4.x-Versionen im Vergleich zu Hive 2.x auf Amazon-EMR-5.x-Versionen verwenden.

ACID-Transaktionen nicht unterstützt

Hive auf Amazon-EMR-4.x-Versionen unterstützt keine ACID-Transaktionen mit Hive-Daten, die in Amazon S3 gespeichert werden, wenn 4.x-Versionen verwendet werden. Wenn Sie versuchen, eine transaktionale Tabelle in Amazon S3 zu erstellen, tritt eine Ausnahme auf.

Lesen und Schreiben von Tabellen in Amazon S3

Hive in Amazon-EMR-4.x-Versionen können direkt ohne die Verwendung temporärer Dateien in Amazon S3 schreiben. Dadurch wird die Leistung verbessert, eine Folge davon ist allerdings, dass Sie nicht in derselben Tabelle in Amazon S3 innerhalb derselben Hive-Anweisung lesen und schreiben können. Sie können dieses Problem umgehen, indem Sie eine temporäre Tabelle in HDFS erstellen und verwenden.

Das folgende Beispiel zeigt, wie Sie mit mehreren Hive-Anweisungen eine Tabelle in Amazon S3 aktualisieren. Die Anweisungen erstellen eine temporäre Tabelle in HDFS mit dem Namen `tmp` auf Grundlage einer Tabelle in Amazon S3 mit dem Namen `my_s3_table`. Die Tabelle in Amazon S3 wird dann mit den Inhalten der temporären Tabelle aktualisiert.

```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;  
INSERT OVERWRITE TABLE tmp SELECT ....;  
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

Log4j im Vergleich zu Log4j 2

Hive auf Amazon-EMR-4.x-Versionen verwendet Log4j. Ab Version 5.0.0, Log4j 2 ist die Standardeinstellung. Diese Versionen erfordern möglicherweise unterschiedliche Protokollierungskonfigurationen. Weitere Informationen finden Sie unter [Apache Log4j 2](#).

MapReduce ist die Standard-Ausführungs-Engine

Hive auf Amazon-EMR-4.x-Versionen verwendet MapReduce als Standard-Ausführungs-Engine. Ab Amazon-EMR-Version 5.0.0 ist Tez die Standardversion, die für die meisten Workflows eine verbesserte Leistung bietet.

Hive-Autorisierung

Hive auf Amazon EMR 4.x-Versionen unterstützt die [Hive-Autorisierung](#) für HDFS, aber nicht für EMRFS und Amazon S3. Amazon-EMR-Cluster werden standardmäßig mit deaktivierter Autorisierung ausgeführt.

Dateizusammenführung mit Hive in Amazon S3

Hive auf Amazon-EMR-4.x-Versionen führt kleine Dateien am Ende eines Map-Auftrags zusammen, wenn `hive.merge.mapfiles true` ist. Eine Zusammenführung wird nur ausgelöst, wenn die durchschnittliche Leistungsgröße des Auftrags kleiner ist als die `hive.merge.smallfiles.avgsize`-Einstellung. Amazon EMR; Hive verhält sich genau gleich, wenn sich der Endausgabepfad in HDFS befindet. Wenn sich der Ausgabepfad in Amazon S3; befindet, wird der Parameter `hive.merge.smallfiles.avgsize` jedoch ignoriert. In diesem Fall wird die Zusammenführungsaufgabe immer dann ausgelöst, wenn `hive.merge.mapfiles` auf `true` gesetzt ist.

Überlegungen für die Verwendung von Pig in Amazon EMR 4.x

Pig-Version 0.14.0 ist auf Clustern installiert, die mit Amazon-EMR-4.x-Versionen erstellt wurden. Für Pig wurde in Amazon EMR 5.0.0 ein Upgrade auf Version 0.16.0 durchgeführt. Die wesentlichen Unterschiede werden unten erläutert.

Unterschiedliche Standard-Ausführungs-Engine

Pig-Version 0.14.0 auf Amazon-EMR-4.x-Versionen verwendet MapReduce als Standard-Ausführungs-Engine. Pig 0.16.0 und höher verwenden Apache Tez. Sie können explizit `execType=mapreduce` in der Konfigurationsklassifizierung `pig-properties` festlegen, um MapReduce zu verwenden.

Gelöschte benutzerdefinierte Pig-Funktionen (UDFs)

Benutzerdefinierte UDFs, die in Pig auf Amazon-EMR-4.x-Versionen waren, waren ab Pig 0.16.0 nicht mehr enthalten. Die meisten UDFs weisen ähnliche Funktionen auf, die Sie stattdessen verwenden können. Die folgende Tabelle listet gelöschte UDFs und ähnliche Funktionen auf. Weitere Informationen finden Sie unter [Integrierte Funktionen](#) auf der Apache-Pig-Website.

Gelöschte UDF	Gleichwertige Funktion
FORMAT_DT(dtformat, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
EXTRACT(string, pattern)	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(date, duration), SubtractDuration(date, duration)
PERIOD(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt,

Gelöschte UDF	Gleichwertige Funktion
	dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
CAPITALIZE(string)	UCFIRST(string)
CONCAT_WITH()	CONCAT()
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()

Die folgenden UDFs wurden ohne gleichwertige Funktionen entfernt: FORMAT(), LOCAL_DATE(), LOCAL_TIME(), CENTER(), LEFT_PAD(), REPEAT(), REPLACE_ONCE(), RIGHT_PAD(), STRIP(), STRIIP_END(), STRIP_START(), SWAP_CASE().

Nicht mehr enthaltene Grunt-Befehle

Einige Grunt-Befehle werden ab Pig 0.16.0 nicht mehr angeboten. In der folgenden Tabelle sind die Grunt-Befehle in Pig 0.14.0 und ihre entsprechenden Befehle (sofern vorhanden) in der aktuellen Version aufgeführt.

Pig 0.14.0 und gleichwertige aktuelle Grunt-Befehle

Grunt-Befehl in Pig 0.14.0	Pig Grunt-Befehl in 0.16.0 und höher
cat <non-hdfs-path>)	fs -cat <non-hdfs-path>;
cd <non-hdfs-path>;	Keine Entsprechung
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>;

Grunt-Befehl in Pig 0.14.0	Pig Grunt-Befehl in 0.16.0 und höher
<code>copyToLocal <non-hdfs-path> <local-path>;</code>	<code>fs -copyToLocal <non-hdfs-path> <local-path>;</code>
<code>copyFromLocal <local-path> <non-hdfs-path>;</code>	<code>fs -copyFromLocal <local-path> <non-hdfs-path>;</code>
<code>mkdir <non-hdfs-path>;</code>	<code>fs -mkdir <non-hdfs-path>;</code>
<code>rm <non-hdfs-path>;</code>	<code>fs -rm -r -skipTrash <non-hdfs-path>;</code>
<code>rmf <non-hdfs-path>;</code>	<code>fs -rm -r -skipTrash <non-hdfs-path>;</code>

Entfernte Funktionen für Nicht-HDFS-Stammverzeichnisse

Pig 0.14.0 in Amazon-EMR-4.x-Versionen verfügt über zwei Mechanismen, um anderen Benutzern als dem `hadoop`-Benutzer, die keine Root-Verzeichnisse besitzen, das Ausführen von Pig-Skripten zu ermöglichen. Der erste Mechanismus ist ein automatisches Fallback, mit dem das anfängliche Arbeitsverzeichnis auf das Stammverzeichnis festgelegt wird, wenn kein Stammverzeichnis vorhanden ist. Der zweite ist eine `pig.initial.fs.name`-Eigenschaft, über die Sie das anfängliche Arbeitsverzeichnis ändern können.

Diese Mechanismen sind beginnend mit Amazon-EMR-Version 5.0.0 nicht verfügbar, und Benutzer müssen ein Stammverzeichnis in HDFS haben. Dies gilt nicht für `hadoop`-Benutzer, da hier ein Stammverzeichnis beim Starten bereitgestellt wird. Bei Skriptläufen, bei denen Hadoop-JAR-Schritte verwendet werden, wird standardmäßig der Hadoop-Benutzer vorgeschlagen, es sei denn, es ist explizit ein anderer Benutzer über `command-runner.jar` angegeben.

EMRAmazon-Version 4.9.6

4.9.6 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
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

Versionshinweise zu 4.9.6

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

4.9.6 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-2	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful-HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.9.6 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.9.6-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.9.5

4.9.5 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
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

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
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	3.4.9

Versionshinweise zu 4.9.5

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 4.9.5. Änderungen beziehen sich auf Version 4.9.4.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- HBase
 - Diese Version behebt eine potenzielle Schwachstellen.

4.9.5 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich

EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	1.0.0-amzn-9	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.9.5 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.9.5-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.9.4

4.9.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
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

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
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

Versionshinweise zu 4.9.4

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 4.9.4. Änderungen beziehen sich auf Version 4.9.3.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert EMR, um potenzielle Sicherheitslücken zu beheben.

4.9.4 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel

mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBaseBefehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	1.0.0-amzn-9	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.9.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.9.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.9.3

4.9.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
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

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
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

Versionshinweise zu 4.9.3

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 4.9.3. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.9.2.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

4.9.3-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle zu HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL-Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Client.

4.9.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.9.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.9.2

4.9.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

4.9.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.9.2. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.9.1.

Veröffentlichungsdatum: 13. Juli 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

4.9.2-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Client.

4.9.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.9.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.

Klassifizierungen	Beschreibung
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraooop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Klassifizierungen	Beschreibung
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.9.1

4.9.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
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

Versionshinweise zu 4.9.1

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.9.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.4.

Veröffentlichungsdatum: 10. April 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- [Backports von HIVE-9976 und -10106 HIVE](#)
- Es wurde ein Problem behoben YARN, bei dem eine große Anzahl von Knoten (mehr als 2.000) und Containern (mehr als 5.000) einen out-of-memory Fehler verursachte, zum Beispiel: "Exception in thread main java.lang.OutOfMemoryError"

Änderungen und Verbesserungen

- EMR Amazon-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- Python 2.6 wurde aus dem EMR Linux-Basisimage von Amazon entfernt. Sie können Python 2.6 bei Bedarf manuell installieren.

4.9.1-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.15.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-2	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-9	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.9.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.9.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.8.5

4.8.5 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
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	0.157.1	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

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
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

Versionshinweise zu 4.8.5

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

4.8.5-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-*

`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
<code>ganglia-metadata-collector</code>	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
<code>ganglia-web</code>	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.3-amzn-1	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-8	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-8	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-8	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-8	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.8.5 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.8.5-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.8.4

4.8.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
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
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

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
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

Versionshinweise zu 4.8.4

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.4. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.3.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

4.8.4-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	1.0.0-amzn-8	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-8	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-8	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.25+	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.8.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.8.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.8.3

4.8.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
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
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

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
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

Versionshinweise zu 4.8.3

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.3. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.2.

Veröffentlichungsdatum: 29. Dezember 2016

Upgrades

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf Spark 1.6.3. Weitere Informationen finden Sie in den [Spark-Versionshinweisen](#) in der Apache Spark-Dokumentation.
- Auf ZooKeeper 3.4.9 aktualisiert. Weitere Informationen finden Sie in den [ZooKeeperVersionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

Änderungen und Verbesserungen

- Unterstützung für den Instance-Typ Amazon EC2 m4.16xlarge in EMR Amazon-Version 4.8.3 und höher hinzugefügt, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem ControlledJob # mit einer Nullzeiger-Ausnahme (NPE) toString fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

4.8.3-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.3-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-8	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	1.0.0-amzn-8	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-8	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilen-Cli-ent.

4.8.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.8.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.8.2

4.8.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
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
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.3

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
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

4.8.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.2. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.0.

Veröffentlichungsdatum: 24. Oktober 2016

Upgrades

- Upgrade auf Hadoop 2.7.3
- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Web Interface](#) in der Presto-Dokumentation.
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

4.8.2-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.10.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	1.0.0-amzn-7	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherv erwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTPEndpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-7	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-7	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfrag en.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderu ngen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.8.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.8.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.8.1

4.8.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
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
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

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.3	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

Versionshinweise zu 4.8.1

4.8.1-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-4	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-4	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.2-amzn-4	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-7	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-7	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-7	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-7	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.51	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.151	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.151	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	HTTPApache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.8.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.8.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Einstellungen ändernEMRFS.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.8.0

4.8.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
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

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
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

Versionshinweise zu 4.8.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.7.2.

Veröffentlichungsdatum: 7. September 2016

Upgrades

- Auf 1.2.2 aktualisiert HBase
- Upgrade auf Presto-Sandbox 0.151
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin-Sandbox 0.6.1

Änderungen und Verbesserungen

- Es wurde ein Problem behoben YARN, bei dem ApplicationMaster versucht wurde, Container zu bereinigen, die nicht mehr existieren, weil ihre Instanzen beendet wurden.
- Die Aktionen hive-server2 URL für Hive2 in den Oozie-Beispielen wurden korrigiert.
- Zusätzliche Presto-Kataloge werden jetzt unterstützt.
- [Rückportierte Patches: HIVE-8948, -12679, -13405, -3116, -12689 HIVE HIVE PHOENIX HADOOP](#)
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

4.8.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.2-amzn-4	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-4	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARNDienst für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.2	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.2	HBase Befehlszeilenclient.
hbase-rest-server	1.2.2	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.2	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-7	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	1.0.0-amzn-7	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-7	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.51	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.151	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.151	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.8.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.8.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.7.4

4.7.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.148	0.147
Spark	1.6.2	1.6.2	1.6.2	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
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

Versionshinweise zu 4.7.4

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen EMR Amazon-Version.

Important

In dieser Release-Version EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

4.7.4 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-*

amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.8.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN Dienst für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.2-amzn-3	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.1	Dienst, der einen RESTful HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-6	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-6	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-6	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-6	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.148	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.148	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.7.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.7.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.7.3

4.7.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
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
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

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
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

4.7.3 Versionshinweise

4.7.3 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.8.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resource-manager	2.7.2-amzn-3	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Dienst für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.1	Dienst für die Versorgung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilenclient.
hbase-rest-server	1.2.1	Dienst, der einen RESTful HTTP-Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-6	Der "hcat"-Befehlszeilenclient für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-6	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-6	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-6	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.148	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.148	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die YARN TEZ-Anwendung und die Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.7.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

EMR-4.7.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Einstellungen ändern EMRFS.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.7.2

4.7.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1.10,27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.0	0.12.0	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	-
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

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

4.7.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR 4.7.2.

Veröffentlichungsdatum: 15. Juli 2016

Features

- Upgrade auf Mahout 0.12.2
- Upgrade auf Presto 0.148
- Upgrade auf Spark 1.6.2
- Sie können jetzt eine AWSCredentialsProvider für die EMRFS Verwendung mit a URI als Parameter erstellen. Weitere Informationen finden Sie unter [Erstellen eines AWSCredentialsProvider für EMRFS](#).

- EMRFS ermöglicht es Benutzern jetzt, mithilfe der `fs.s3.consistent.dynamodb.endpoint` Eigenschaft in einen benutzerdefinierten DynamoDB-Endpunkt für ihre Consistent View-Metadaten zu konfigurieren. `emrfs-site.xml`
- Es wurde ein Skript in `/usr/bin` mit dem Namen `spark-example` hinzugefügt, das `/usr/lib/spark/spark/bin/run-example` umschließt, sodass Sie Beispiele direkt ausführen können. Um beispielsweise das SparkPi Beispiel auszuführen, das in der Spark-Distribution enthalten ist, können Sie es `spark-example SparkPi 100` von der Befehlszeile aus ausführen oder `command-runner.jar` als Schritt in der API

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem Oozie die Datei `spark-assembly.jar` nicht am richtigen Speicherort ablegte, wenn Spark auch installiert war. Dies führte dazu, dass Spark-Anwendungen nicht mit Oozie gestartet werden konnten.
- Ein Problem mit der auf Spark Log4J basierenden Protokollierung in Containern wurde behoben. YARN

4.7.2-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.8.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basierten KeyProvider API basiert.
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.2-amzn-3	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.2-amzn-3	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.1	HBase Befehlszeilenclient.
hbase-rest-server	1.2.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-6	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.

Komponente	Version	Beschreibung
hive-client	1.0.0-amzn-6	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-6	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.148	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.148	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.23	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.7.2 Klassifizierungen der Konfiguration

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.7.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.7.1

4.7.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
AWS SDK for Java	1.10.75	1,10,75	1.10,27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.0	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.12.0	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	-	-
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

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
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

4.7.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR 4.7.1.

Veröffentlichungsdatum: 10. Juni 2016

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, das die Startzeit von Clustern verlängerte, die in einem VPC mit privaten Subnetzen gestartet wurden. Der Fehler betraf nur Cluster, die mit der EMR Amazon-Version 4.7.0 gestartet wurden.
- Es wurde ein Problem behoben, bei dem das Auflisten von Dateien in Amazon EMR für Cluster, die mit der Version Amazon EMR 4.7.0 gestartet wurden, nicht richtig behandelt wurde.

4.7.1-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.7.1	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-2	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API

Komponente	Version	Beschreibung
hadoop-mapred	2.7.2-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.1	HBase Befehlszeilenclient.
hbase-rest-server	1.2.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	1.0.0-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-5	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP-Endpunkt, der eine REST-Schnittstelle zu HCatalog bereitstellt.
hive-client	1.0.0-amzn-5	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-5	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL-Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.0	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API
presto-coordinator	0.147	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.147	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungseengine für. YARN
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.18	HTTPApache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Client.

4.7.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml` Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.7.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hbase-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Werte in der Umgebung ebHCCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.7.0

4.7.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
AWS SDK for Java	1.10.75	1.10,27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
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
Phoenix	4.7.0	-	-	-

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
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	-

Versionshinweise zu 4.7.0

Important

Amazon EMR 4.7.0 ist veraltet. Verwenden Sie stattdessen Amazon EMR 4.7.1 oder höher.

Veröffentlichungsdatum: 2. Juni 2016

Features

- Apache Phoenix 4.7.0 wurde hinzugefügt.
- Apache Tez 0.8.3 wurde hinzugefügt.
- Auf 1.2.1 aktualisiert HBase
- Upgrade auf Mahout 0.12.0
- Upgrade auf Presto 0.147
- Auf 1.10.75 AWS SDK for Java aktualisiert
- Das letzte Flag wurde von der Eigenschaft `mapreduce.cluster.local.dir` in `mapred-site.xml` entfernt, damit Benutzer Pig im lokalen Modus ausführen können.
- Amazon Redshift JDBC Redshift-Treiber im Cluster verfügbar

Amazon Redshift JDBC Redshift-Treiber sind jetzt unter `/usr/share/aws/redshift/jdbc` enthalten. `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` ist der JDBC 4.1-kompatible Amazon Redshift Redshift-Treiber und `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` ist der JDBC 4.0-kompatible Amazon Redshift Redshift-Treiber. Weitere Informationen finden [Sie unter Configure a JDBC Connection](#) im Amazon Redshift Management Guide.

- Java 8

Mit Ausnahme von Presto ist Open JDK 1.7 der Standard, der für alle JDK Anwendungen verwendet wird. Es sind jedoch sowohl Open JDK 1.7 als auch 1.8 installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Kernelproblem behoben, das die Leistung von durchsatzoptimierten HDD (st1) EBS Volumes für Amazon EMR in `emr-4.6.0` erheblich beeinträchtigte.
- Es wurde ein Problem behoben, bei dem ein Cluster fehlschlug, wenn eine HDFS Verschlüsselungszone angegeben wurde, ohne Hadoop als Anwendung auszuwählen.
- Die HDFS Standard-Schreibrichtlinie wurde von `RoundRobin` zu `AvailableSpaceVolumeChoosingPolicy` geändert. Einige Volumes wurden mit der `RoundRobin` Konfiguration nicht richtig genutzt, was zu einem Ausfall der Core-Knoten und zu einem unzuverlässigen Betrieb HDFS führte.

- Es wurde ein Problem mit dem behoben EMRFSCLI, das bei der Erstellung der standardmäßigen DynamoDB-Metadatentabelle für konsistente Ansichten zu einer Ausnahme führte.
- Es wurde ein Deadlock-Problem behoben EMRFS, das möglicherweise bei mehrteiligen Umbenennungs- und Kopiervorgängen auftrat.
- Es wurde ein Problem behoben EMRFS, das dazu führte, dass die CopyPart Standardgröße 5 MB betrug. Der Standard ist jetzt ordnungsgemäß auf 128 MB eingestellt.
- Es wurde ein Problem mit der Zeppelin-Startkonfiguration behoben, das möglicherweise das Anhalten des Service verhinderte.
- Es wurde ein Problem mit Spark und Zeppelin behoben, das Sie daran hinderte, das `s3a://` URI Schema zu verwenden, weil es nicht richtig in den jeweiligen Klassenpfad geladen `/usr/lib/hadoop/hadoop-aws.jar` wurde.
- [Rückportiert -2484. HUE](#)
- Ein [Commit](#) von Hue 3.9.0 (JIRA existiert nicht) wurde zurückportiert, um ein Problem mit dem Browser-Beispiel zu beheben. HBase
- [Rückportiert -9073HIVE.](#)

4.7.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
emr-ddb	3.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.7.1	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-2	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN-Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.2-amzn-2	YARN-Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.2-amzn-2	Dienst zum Abrufen aktueller und historischer Informationen für YARN Anwendungen.
hbase-hmaster	1.2.1	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen zuständig ist.
hbase-region-server	1.2.1	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.1	HBase Befehlszeilenclient.
hbase-rest-server	1.2.1	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.1	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-5	Bereitstellung von Diensten HCatalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.

Komponente	Version	Beschreibung
hive-client	1.0.0-amzn-5	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-5	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.0	Bibliothek für Machine Learning.
mysql-server	5.5.46	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0- -1.2 HBase	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4,7,0- -1,2 HBase	Ein leichter Server, der sowohl JDBC Zugriff als auch Protokollpuffer und JSON Formatzugriff auf die Avatica bietet API

Komponente	Version	Beschreibung
presto-coordinator	0.147	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.147	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungseengine für YARN
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die YARN TEZ-Anwendung und die Bibliotheken.
webserver	2.4.18	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Cli-ent.

4.7.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.7.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.

Klassifizierungen	Beschreibung
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.

Klassifizierungen	Beschreibung
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.6.1

4.6.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#), und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	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	-	-	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
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	-	-

4.6.1 Versionshinweise

4.6.1 Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.3.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.6.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.2-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen verantwortlich ist.
hbase-region-server	1.2.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.0	HBase Befehlszeilenclient.
hbase-rest-server	1.2.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-4	Bereitstellung von Diensten H Catalog, einer Tabelle und einer Speicherebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	1.0.0-amzn-4	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-6	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.143	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.143	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Client.

4.6.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.6.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Einstellungen ändern EMRFS.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Ändern Sie die Werte in der Datei <code>hadoop-metrics2-hbase.properties</code> . HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei <code>hbase-policy.xml</code> .
hbase-site	Ändern Sie die Werte in HBase der Datei <code>hbase-site.xml</code> .
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.

Klassifizierungen	Beschreibung
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.6.0

4.6.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#), und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	-	-	-
HCatalog	1.0.0	1.0.0	1.0.0	-
Hadoop	2.7.2	2.7.2	2.7.1	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
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

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	-	-	-

Versionshinweise zu 4.6.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.6.0.

- 1.2.0 wurde hinzugefügt HBase
- ZooKeeper-Sandbox 3.4.8 wurde hinzugefügt.
- Upgrade auf Presto-Sandbox 0.143
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.03.0. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.
- Problem mit Auswirkungen auf durchsatzoptimierte HDD (st1) Volumetypen EBS

Ein Problem in den Linux-Kernel-Versionen 4.2 und höher beeinträchtigt die Leistung von durchsatzoptimierten HDD (st1) EBS Volumes für erheblich. EMR Diese Version (emr-4.6.0) verwendet die Kernel-Version 4.4.5 und ist daher betroffen. Daher empfehlen wir, emr-4.6.0 nicht zu verwenden, wenn Sie st1-Volumes verwenden möchten. EBS Sie können emr-4.5.0 oder

frühere EMR Amazon-Versionen mit st1 ohne Auswirkungen verwenden. Außerdem stellen wir die Problembhebung mit späteren Versionen zur Verfügung.

- Python-StandardEinstellungen

Python 3.4 ist jetzt standardmäßig installiert, Python 2.7 bleibt jedoch der Systemstandard. Sie können Python 3.4 mit einer Bootstrap-Aktion als Systemstandard konfigurieren. Sie können die Konfiguration verwenden, API um `PYSPARK_PYTHON export /usr/bin/python3.4` in der `spark-env` Klassifizierung festzulegen, um die Python-Version zu beeinflussen, die von PySpark verwendet wird.

- Java 8

Mit Ausnahme von Presto ist Open JDK 1.7 der Standard, der für alle JDK Anwendungen verwendet wird. Es sind jedoch sowohl Open JDK 1.7 als auch 1.8 installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die Anwendungsbereitstellung manchmal wegen eines generierten Passwort nach dem Zufallsprinzip ausfiel.
- Bisher war `mysql` auf allen Knoten installiert. Jetzt ist es nur auf der Master-Instance installiert und nur dann, wenn die gewählte Anwendung `mysql-server` als Komponente enthält. Derzeit enthalten die folgenden Anwendungen die `mysql-server` Komponente: Hive, HueHCatalog, Presto-Sandbox und Sqoop-Sandbox.
- Von der Standardeinstellung `32 yarn.scheduler.maximum-allocation-vcores` auf 80 geändert, wodurch ein in `emr-4.4.0` eingeführtes Problem behoben wird, das hauptsächlich bei Spark auftritt, wenn die `maximizeResourceAllocation` Option in einem Cluster verwendet wird, dessen Core-Instance-Typ einer von wenigen großen Instance-Typen ist, bei denen die YARN `vcores` höher als 32 eingestellt sind; nämlich `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` oder `m4.10xlarge` wurden von diesem Problem betroffen.
- `s3-dist-cp` verwendet jetzt EMRFS für alle Amazon S3 S3-Nominierungen und speichert nicht mehr in einem temporären Verzeichnis. HDFS
- Es wurde ein Problem mit der Ausnahmebehandlung für die clientseitigen Verschlüsselung mehrteiliger Uploads behoben.

- Es wurde eine Option hinzugefügt, mit der Benutzer die Amazon-S3-Speicherklassen ändern können. Standardmäßig ist diese Einstellung STANDARD. Die Konfigurationsklassifizierungseinstellung `emrfs-site` ist `fs.s3.storageClass` und die möglichen Werte sind STANDARD, STANDARD_IA und REDUCED_REDUNDANCY. Weitere Informationen zu Speicherklassen finden Sie unter [Speicherklassen](#) im Benutzerhandbuch zu Amazon Simple Storage Service.

4.6.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.3.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.6.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-1	HDFS-Befehlszeilenclient und Bibliothek

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httfs-server	2.7.2-amzn-1	HTTPEndpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hbase-hmaster	1.2.0	Dienst für einen HBase Cluster, der für die Koordination der Regionen und die Ausführung von Verwaltungsbefehlen verantwortlich ist.
hbase-region-server	1.2.0	Dienst für die Versorgung einer oder mehrerer HBase Regionen.
hbase-client	1.2.0	HBaseBefehlszeilenclient.

Komponente	Version	Beschreibung
hbase-rest-server	1.2.0	Dienst, der einen RESTful HTTP Endpunkt für bereitstellt. HBase
hbase-thrift-server	1.2.0	Dienst, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-4	Bereitstellung von DienstenH Catalog, einer Tabelle und einer Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstelltHCatalog.
hive-client	1.0.0-amzn-4	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-6	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
mahout-client	0.11.1	Bibliothek für Machine Learning.
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.143	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.143	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungseengine fürYARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilen-Client.

4.6.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.6.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hbase-env	Werte in HBase der Umgebung ändern.
hbase-log4j	Ändern Sie die Werte in der Datei HBase hbase-log4j.properties.
hbase-metrics	Ändern Sie die Werte in der Datei hadoop-metrics2-hbase.properties. HBase
hbase-policy	Ändern Sie HBase die Werte in der Datei hbase-policy.xml.
hbase-site	Ändern Sie die Werte in HBase der Datei hbase-site.xml.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's jndi.properties.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

EMRAmazon-Version 4.5.0

4.5.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
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)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.5	0,5,5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Versionshinweise zu 4.5.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.5.0.

Veröffentlichungsdatum: 4. April 2016

Features

- Upgrade auf Spark 1.6.1
- Upgrade auf Hadoop 2.7.2
- Upgrade auf Presto 0.140
- AWS KMS Unterstützung für serverseitige Amazon S3 S3-Verschlüsselung hinzugefügt.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem My SQL - und Apache-Server nach dem Neustart eines Knotens nicht gestartet wurden.
- Es wurde ein Problem behoben, das mit IMPORT nicht partitionierten Tabellen, die in Amazon S3 gespeichert waren, nicht richtig funktionierte
- Es wurde ein Problem mit Presto behoben, bei dem das Staging-Verzeichnis beim Schreiben in Hive-Tabellen /mnt/tmp anstatt /tmp lauten muss.

4.5.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.2.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.5.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-0	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.2-amzn-0	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.2-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.2-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.2-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hcatalog-client	1.0.0-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-4	Bereitstellung von Diensten H Catalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle zu bereitstellt H Catalog.
hive-client	1.0.0-amzn-4	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-4	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.140	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.140	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	HTTP Apache-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

4.5.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.5.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's <code>.xmlHCatalog.proto-hive-site</code>
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

EMRAmazon-Version 4.4.0

4.4.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-
HCatalog	1.0.0	-	-	-
Hadoop	2.7.1	2.7.1	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
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	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.5	0,5,5	0,6,0-SNAPSHOT
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Versionshinweise zu 4.4.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.4.0.

Veröffentlichungsdatum: 14. März 2016

Features

- 1.0.0 wurde hinzugefügt HCatalog
- Sqoop-Sandbox 1.4.6 wurde hinzugefügt.
- Upgrade auf Presto 0.136
- Upgrade auf Zeppelin 0.5.6
- Upgrade auf Mahout 0.11.1
- `dynamicResourceAllocation` wurde standardmäßig aktiviert.
- Es wurde eine Tabelle mit allen Konfigurationsklassifizierungen für die Version hinzugefügt. Weitere Informationen finden Sie in der Konfigurationsklassifizierungstabelle unter [Konfigurieren von Anwendungen](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die `maximizeResourceAllocation` Einstellung nicht genug Speicher für YARN ApplicationMaster Daemons reservierte.
- Es wurde ein Problem behoben, das bei einem benutzerdefinierten Gerät aufgetreten ist. DNS Wenn Einträge in `resolve.conf` den angegebenen, benutzerdefinierten Einträge vorangestellt werden, können die benutzerdefinierten Einträge nicht aufgelöst werden. Dieses Verhalten wurde durch Cluster in einem beeinflusst, in VPC dem der VPC Standard-Nameserver als oberster Eintrag eingefügt wurde `resolve.conf`.
- Es wurde ein Problem behoben, bei dem die Python-Standardversion auf Version 2.7 geändert wurde und boto für diese Version nicht installiert war.
- Es wurde ein Problem behoben, bei dem YARN Container und Spark-Anwendungen eine eindeutige Ganglia-Round-Robin-Datenbankdatei (RRD) generierten, was dazu führte, dass die erste Festplatte, die an die Instanz angeschlossen war, voll wurde. Aufgrund dieses Fixes wurden Metriken auf YARN Containerebene und Kennzahlen auf Spark-Anwendungsebene deaktiviert.
- Es wurde ein Problem im Protokoll-Pusher behoben, bei dem alle leeren Protokollordner gelöscht wurden. Das hatte zur Folge, dass Hive CLI nicht protokollieren konnte, weil der Log-Pusher den leeren `user` Ordner darunter entfernte. `/var/log/hive`

- Es wurde ein Problem mit Hive-Importen behoben, da sich auf die Partitionierung auswirkte und beim Import zu einem Fehler führte.
- Es wurde ein Problem behoben, bei dem EMRFS s3-dist-cp Bucket-Namen, die Punkte enthielten, nicht richtig behandelte.
- Das Verhalten wurde EMRFS dahingehend geändert, dass in Buckets mit aktivierter Versionierung die `_$folder$` Markierungsdatei nicht kontinuierlich erstellt wird, was zu einer verbesserten Leistung für Buckets mit Versionierung beitragen kann.
- Das Verhalten wurde EMRFS dahingehend geändert, dass es keine Befehlsdateien verwendet, außer in Fällen, in denen die clientseitige Verschlüsselung aktiviert ist. Wenn Sie Anweisungsdateien bei Verwendung der clientseitigen Verschlüsselung löschen möchten, können Sie die `emrfs-site.xml`-Eigenschaft, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, auf "true" festlegen.
- Die YARN Protokollaggregation wurde dahingehend geändert, dass die Protokolle zwei Tage lang am Aggregationsziel aufbewahrt werden. Das Standardziel ist Ihr HDFS Clusterspeicher. Wenn Sie diese Dauer ändern möchten, ändern Sie den Wert `yarn.log-aggregation.retain-seconds` mit der Konfigurationsklassifizierung `yarn-site` beim Erstellen Ihres Clusters. Wie immer können Sie Ihre Anwendungsprotokolle beim Erstellen Ihres Clusters mit dem Parameter `log-uri` in Amazon S3 speichern.

Angewendete Patches

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

4.4.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.2.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.4.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.1-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.1-amzn-1	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.1-amzn-1	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.1-amzn-1	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-https-server	2.7.1-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.1-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.1-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.1-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.1-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hcatalog-client	1.0.0-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-3	Bereitstellung von Diensten HCatalog, eine Tabellen- und Speicherverwaltungsebene für verteilte Anwendungen.
hcatalog-webhcat-server	1.0.0-amzn-3	HTTP Endpunkt, der eine REST Schnittstelle zu bereitstellt HCatalog.
hive-client	1.0.0-amzn-3	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-3	Dienst für den Zugriff auf den Hive-Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.136	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.136	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.0	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.0	In-Memory-Ausführungseengine fürYARN.
spark-yarn-slave	1.6.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

4.4.0-Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.4.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hcatalog-env	Werte in HCatalog der Umgebung ändern.
hcatalog-server-jndi	Ändern Sie die Werte in HCatalog's <code>jndi.properties</code> .

Klassifizierungen	Beschreibung
hcatalog-server-proto-hive-site	Ändern Sie die Werte in's .xmlHCatalog. proto-hive-site
hcatalog-webhcat-env	Werte in der Umgebung ebHCat von HCatalog W ändern.
hcatalog-webhcat-log4j	Ändern Sie die Werte in den ebHCat log4j.properties von HCatalog W.
hcatalog-webhcat-site	Ändern Sie die Werte in der Datei HCatalog webhcat-site.xml ebHCat von W.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oracoop-site	Ändern Sie die Werte in der Datei oracoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Werte in der YARN Umgebung ändern.
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

EMRAmazon-Version 4.3.0

4.3.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0,6,0-SNAPSHOT	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Versionshinweise zu 4.3.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.3.0.

Veröffentlichungsdatum: 19. Januar 2016

Features

- Upgrade auf Hadoop 2.7.1
- Upgrade auf Spark 1.6.0
- Upgrade auf Ganglia 3.7.2
- Upgrade auf Presto 0.130
- Amazon EMR hat einige Änderungen daran vorgenommen, `spark.dynamicAllocation.enabled` wann es auf `true` gesetzt ist; es ist standardmäßig falsch. Wenn die Einstellung auf `"true"` festgelegt ist, wirkt sich dies auf von der Einstellung `maximizeResourceAllocation` vorgenommene Standardeinstellungen aus:
 - Wenn `spark.dynamicAllocation.enabled` auf `"true"` eingestellt ist, wird `spark.executor.instances` nicht von `maximizeResourceAllocation` festgelegt.
 - Die Einstellung `spark.driver.memory` wird nun basierend auf den Instance-Typen im Cluster auf ähnliche Weise konfiguriert, wie `spark.executors.memory` festgelegt wird. Da die Spark-Treiberanwendung jedoch entweder auf der Master- oder auf einer der Core-Instances ausgeführt werden kann (z. B. im YARN Client- bzw. Cluster-Modus), wird die `spark.driver.memory` Einstellung auf der Grundlage des Instance-Typs des kleineren Instance-Typs zwischen diesen beiden Instance-Gruppen festgelegt.
 - Die `spark.default.parallelism` Einstellung ist jetzt auf die doppelte Anzahl der für YARN Container verfügbaren CPU Kerne festgelegt. In früheren Versionen war dies die Hälfte des Werts.

- Die Berechnungen für den für YARN Spark-Prozesse reservierten Speicher-Overhead wurden angepasst, um genauer zu sein, was zu einer geringfügigen Erhöhung der Gesamtspeichermenge führte, die Spark zur Verfügung steht (d. h. `spark.executor.memory`).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- YARN Die Protokollaggregation ist jetzt standardmäßig aktiviert.
- Es wurde ein Problem behoben, bei dem Protokolle nicht in den Amazon S3 S3-Log-Bucket für den Cluster übertragen wurden, wenn die YARN Protokollaggregation aktiviert war.
- YARN Die Containergrößen haben jetzt ein neues Minimum von 32 für alle Knotentypen.
- Es wurde ein Problem mit Ganglia behoben, dass zu übermäßigen Festplatten-E/A-Vorgängen auf dem Primärknoten in großen Clustern führte.
- Es wurde ein Problem behoben, das verhinderte, dass Anwendungsprotokolle per Push-Verfahren an Amazon S3 übertragen wurden, wenn ein Cluster heruntergefahren wird.
- Es wurde ein Problem behoben EMRFSCCLI, das dazu führte, dass bestimmte Befehle fehlschlügen.
- Es wurde ein Problem mit Zeppelin behoben, das verhinderte, dass Abhängigkeiten in die Basisdatei geladen wurden. SparkContext
- Es wurde ein Problem behoben, das aus einer Größenanpassung resultierte bei dem Versuch, Instances hinzuzufügen.
- Es wurde ein Problem in Hive behoben, bei dem CREATE TABLE AS übermäßig SELECT viele Listenaufrufe an Amazon S3 tätigte.
- Es wurde ein Problem behoben, bei dem große Cluster nicht ordnungsgemäß bereitgestellt wurden wenn Hue, Oozie und Ganglia installiert sind.
- Es wurde ein Problem in s3-dist-cp behoben, bei dem ein Beendigungscode gleich Null zurückgegeben wurde, wenn ein Fehler auftrat.

4.3.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind

normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die myapp-component mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.1.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.3.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.1-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.1-amzn-0	HDFS Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.1-amzn-0	HDFS Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.7.1-amzn-0	HDFS Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.7.1-amzn-0	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.7.1-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.7.1-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.1-amzn-0	YARNDienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.1-amzn-0	YARNDienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-2	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-2	Dienst für den Zugriff auf den Hive Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.0	Bibliothek für Machine Learning.
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
presto-coordinator	0.130	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.130	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.0	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.0	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	1.6.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
webserver	2.4	HTTP Apache-Server.
zeppelin-server	0.5.5-incubating-amzn-1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

4.3.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.3.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	EMRFSEinstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's hdfs-site.xml.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei hive-exec-log4j.properties von Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
yarn-env	Werte in der Umgebung ändern. YARN
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

EMRAmazon-Version 4.2.0

4.2.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0,6,0-SNAPSHOT	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Versionshinweise zu 4.2.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.2.0.

Veröffentlichungsdatum: 18. November 2015

Features

- Ganglia-Support wurde hinzugefügt.
- Upgrade auf Spark 1.5.2
- Upgrade auf Presto 0.125
- Upgrade auf Oozie 4.2.0
- Upgrade auf Zeppelin 0.5.5
- Auf 1.10.27 AWS SDK for Java aktualisiert

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben EMRFSCLI, bei dem der Standardname der Metadatentabelle nicht verwendet wurde.
- Es wurde ein Problem behoben, das bei der Verwendung von Tabellen mit ORC -back in Amazon S3 auftrat.
- Es wurde ein Problem behoben, das bei einem Python-Versionskonflikt in der Spark-Konfiguration auftrat.
- Es wurde ein Problem behoben, bei dem der Status eines YARN Knotens aufgrund von DNS Problemen mit Clustern in einem VPC nicht gemeldet wurde.
- Es wurde ein Problem behoben, das YARN auftrat, wenn Knoten außer Betrieb genommen wurden, was dazu führte, dass Anwendungen hängen blieben oder neue Anwendungen nicht geplant werden konnten.

- Es wurde ein Problem behoben, das auftrat, wenn Cluster mit dem Status TIMED _ OUT _ STARTING beendet wurden.
- Es wurde ein Problem behoben, das auftrat, wenn die EMRFS Scala-Abhängigkeit in andere Builds aufgenommen wurde. Die Scala-Abhängigkeit wurde entfernt.

4.2.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.0.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emrfs	2.2.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.6.0	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.6.0	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.5.10	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.6.0-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.6.0-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.6.0-amzn-2	HDFS-Befehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.6.0-amzn-2	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.6.0-amzn-2	HTTP-Endpoint für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.6.0-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.6.0-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.6.0-amzn-2	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.6.0-amzn-2	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-1	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-1	Dienst für den Zugriff auf den Hive Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.125	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.125	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.5.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.5.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.5.2	In-Memory-Ausführungseengine fürYARN.
spark-yarn-slave	1.5.2	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.
webserver	2.4	HTTPApache-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.5.5-incubating-amzn-0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

4.2.0-Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.2.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei <code>hive-exec-log4j.properties</code> von Hive.

Klassifizierungen	Beschreibung
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
yarn-env	Werte in der Umgebung ändern. YARN
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

EMRAmazon-Version 4.1.0

4.1.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0,6,0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Versionshinweise zu 4.1.0

4.1.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert

wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt 2.2-amzn-2.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.0.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.1.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
hadoop-client	2.6.0-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.6.0-amzn-1	HDFSService auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.6.0-amzn-1	HDFSBefehlszeilenclient und Bibliothek
hadoop-hdfs-namenode	2.6.0-amzn-1	HDFSDienst zum Verfolgen von Dateinamen und Blockspeicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.6.0-amzn-1	HTTP Endpunkt für HDFS Operationen.
hadoop-kms-server	2.6.0-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf dem von Hadoop basiert. KeyProvider API
hadoop-mapred	2.6.0-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.6.0-amzn-1	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcenmanager	2.6.0-amzn-1	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-1	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-1	Dienst für den Zugriff auf den Hive Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-4	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
mahout-client	0.11.0	Bibliothek für Machine Learning.
mysql-server	5.5	Mein Datenbankserver. SQL
oozie-client	4.0.1	Oozie-Befehlszeilen-Client.
oozie-server	4.0.1	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.119	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.119	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.5.0	Spark-Befehlszeilen-Clients.
spark-history-server	1.5.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.5.0	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	1.5.0	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

Komponente	Version	Beschreibung
zeppelin-server	0.6.0-inkubieren- SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

4.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfigurieren Sie HDFS Verschlüsselungszonen.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei <code>hive-exec-log4j.properties</code> von Hive.

Klassifizierungen	Beschreibung
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Werte in der KMS Hadoop-Umgebung ändern.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
yarn-env	Werte in der Umgebung ändern. YARN
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

EMRAmazon-Version 4.0.0

4.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Hadoop](#), [Hive](#), [Mahout](#), [Pig](#) und [Spark](#).

In der folgenden Tabelle sind die in dieser Version von Amazon verfügbaren Anwendungsversionen EMR und die Anwendungsversionen der vorherigen drei EMR Amazon-Versionen (sofern zutreffend) aufgeführt.

Eine umfassende Historie der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon EMR 7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
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

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0,6,0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Versionshinweise zu 4.0.0

4.0.0-Komponentenversionen

Die Komponenten, die Amazon mit dieser Version EMR installiert, sind unten aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur bei Amazon erhältlich EMR und werden für Systemprozesse und Funktionen installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der neuesten EMR Amazon-Version sind normalerweise die neueste Version, die in der Community zu finden ist. Wir stellen Community-Veröffentlichungen EMR so schnell wie möglich auf Amazon zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn beispielsweise eine Open-Source-Community-Komponente, die `myapp-component` mit Version 2.2 benannt ist, dreimal geändert wurde, um sie in verschiedenen EMR Amazon-Versionen aufzunehmen, wird ihre Release-Version als aufgeführt `2.2-amzn-2`.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.0.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.0.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.0.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
hadoop-client	2.6.0-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.6.0-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-namenode	2.6.0-amzn-0	HDFS-Dienst zum Verfolgen von Dateinamen und Blockspeicherorten.
hadoop-httpfs-server	2.6.0-amzn-0	HTTP-Endpoint für HDFS-Operationen.
hadoop-mapred	2.6.0-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce-Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.6.0-amzn-0	YARN Dienst zur Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.6.0-amzn-0	YARN Dienst für die Zuweisung und Verwaltung von Clusterressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-0	Dienst für den Zugriff auf den Hive Metastore, ein semantisches Repository, das Metadaten für SQL Hadoop-Operationen speichert.
hive-server	1.0.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
mahout-client	0.10.0	Bibliothek für Machine Learning.
mysql-server	5.5	Mein Datenbankserver. SQL
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.4.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.4.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	1.4.1	In-Memory-Ausführungseengine für YARN.
spark-yarn-slave	1.4.1	Apache Spark-Bibliotheken, die von YARN Slaves benötigt werden.

4.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML Konfigurationsdatei für die Anwendung, wie z. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

emr-4.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	EMRFS-Einstellungen ändern.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-site	Ändern Sie die Werte in HDFS's <code>hdfs-site.xml</code> .
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei <code>hive-exec-log4j.properties</code> von Hive.

Klassifizierungen	Beschreibung
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
httpfs-env	Werte in der Umgebung ändern. HTTPFS
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
mapred-env	Werte in der MapReduce Anwendung sumgebung ändern.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
spark	Amazon EMR — kuratierte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
yarn-env	Werte in der Umgebung ändern. YARN
yarn-site	Ändern Sie die Werte in YARN der Datei yarn-site.xml.

Amazon EMR 2.x- und AMI 3.x-Versionen

Note

AWS aktualisiert die TLS Konfiguration für alle AWS API Endgeräte auf eine Mindestversion von 1.2. TLS EMR Amazon-Versionen 3.10 und niedriger unterstützen nur TLS 1.0/1.1-Verbindungen. Nach dem 4. Dezember 2023 können Sie mit Amazon EMR 3.10 und niedriger keine Cluster mehr erstellen.

Wenn Sie Amazon EMR 3.10 oder niedriger verwenden, empfehlen wir Ihnen, Ihre Workloads sofort zu testen und auf die neueste EMR Amazon-Version zu migrieren. Weitere Informationen erhalten Sie im [AWS -Sicherheitsblog](#).

Die Versionen EMR 2.x und 3.x von Amazon, sogenannte AMI Versionen, werden für bereits bestehende Lösungen zur Verfügung gestellt, die sie aus Kompatibilitätsgründen benötigen. Sie sollten mit diesen Versionen keine neuen Cluster oder Lösungen erstellen. In diesen Versionen fehlen Funktionen, die in neueren Versionen vorhanden sind. Zudem enthalten sie veraltete Anwendungspakete.

Wir empfehlen Ihnen, Lösungen mit der neuesten EMR Amazon-Release-Version zu erstellen.

Der Umfang der Unterschiede zwischen den Release-Versionen der Serien 2.x und 3.x und den neuesten EMR Amazon-Release-Versionen ist erheblich. Der Umfang der Unterschiede reicht von der Erstellung und Konfiguration eines Clusters bis hin zur Port- und Verzeichnisstruktur von Anwendungen auf dem Cluster.

In diesem Abschnitt wird versucht, die wichtigsten Unterschiede für Amazon EMR sowie spezifische Unterschiede bei der Konfiguration und Verwaltung von Anwendungen zu behandeln. Die Angaben sind nicht vollständig. Wenn Sie Cluster mit den Versionen 2.x oder 3.x erstellen und nutzen, stoßen Sie möglicherweise auf Unterschiede, die in diesem Abschnitt nicht aufgeführt sind.

Themen

- [Erstellen eines Clusters mit früheren AMI Versionen von Amazon EMR](#)
- [Installation von Anwendungen mit früheren AMI Versionen von Amazon EMR](#)
- [Anpassen der Cluster- und Anwendungskonfiguration mit früheren AMI Versionen von Amazon EMR](#)
- [Hive-Anwendungsspezifikationen für frühere AMI Versionen von Amazon EMR](#)

- [HBaseAnwendungsspezifikationen für frühere AMI Versionen von Amazon EMR](#)
- [Pig-Anwendungsspezifikationen für frühere AMI Versionen von Amazon EMR](#)
- [Spark-Anwendungsspezifikationen mit früheren AMI Versionen von Amazon EMR](#)
- [Unterschiede zwischen dem DistCp S3-Hilfsprogramm und früheren AMI Versionen von Amazon EMR](#)

Erstellen eines Clusters mit früheren AMI Versionen von Amazon EMR

Die Versionen EMR 2.x und 3.x von Amazon werden je nach Version referenziert AMI. Bei EMR Amazon-Version 4.0.0 und höher werden Releases nach Release-Version referenziert, wobei ein Release-Label wie `emr-5.11.0` verwendet wird. Diese Änderung ist am deutlichsten, wenn Sie einen Cluster programmgesteuert mithilfe von AWS CLI oder erstellen.

Wenn Sie den verwenden AWS CLI , um einen Cluster mithilfe einer AMI Release-Version zu erstellen, verwenden Sie die `--ami-version` Option, z. B. `--ami-version 3.11.0` Viele Optionen, Funktionen und Anwendungen, die in Amazon EMR 4.0.0 und höher eingeführt wurden, sind nicht verfügbar, wenn Sie eine `--ami-version` angeben. Weitere Informationen finden Sie unter [create-cluster](#) in der AWS CLI -Befehlsreferenz.

Der folgende AWS CLI Beispielbefehl startet einen Cluster mit einer AMI Version.

Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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"]
```

Programmgesteuert verwenden alle EMR Amazon-Release-Versionen die `RunJobFlowRequest` Aktion in der, EMR API um Cluster zu erstellen. Der folgende Java-Beispielcode erstellt einen Cluster mit der AMI Release-Version 3.11.0.

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("AmiVersion Cluster")
    .withAmiVersion("3.11.0")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge");
```

Der folgende `RunJobFlowRequest`-Aufruf verwendet stattdessen eine Versionsbezeichnung:

```
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");
```

Konfigurieren der Cluster-Größe

Wenn Ihr Cluster ausgeführt wird, bestimmt Hadoop die Anzahl der zum Verarbeiten der Daten erforderlichen Mapper- und Reducer-Tasks. Große Cluster sollten für eine bessere Ressourcennutzung und für kürzere Verarbeitungszeiten über mehr Tasks verfügen. In der Regel bleibt ein EMR Cluster während des gesamten Clusters gleich groß. Sie legen die Anzahl der Aufgaben fest, wenn Sie den Cluster erstellen. Wenn Sie die Größe eines laufenden Clusters ändern, können Sie die Verarbeitung während der Cluster-Ausführung anpassen. Statt mit einer festen Anzahl von Tasks zu arbeiten, können Sie die Anzahl der Tasks während der Nutzungsdauer des Clusters anpassen. Es gibt zwei Konfigurationsoptionen, mit denen die ideale Anzahl von Tasks festgelegt werden kann:

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

Sie können beide Optionen über die `mapred-conf.xml`-Datei festlegen. Wenn Sie einen Auftrag an den Cluster senden, überprüft der Auftrags-Client die Gesamtanzahl der aktuell im gesamten Cluster verfügbaren Map- und Reduce-Slots. Der Auftrags-Client verwendet die folgenden Gleichungen, um die Anzahl der Tasks festzulegen:

- `mapred.map.tasks = mapred.map.tasksperslot * Map-Slots im Cluster`
- `mapred.reduce.tasks = mapred.reduce.tasksperslot * Reduce-Slots im Cluster`

Wenn die Anzahl der Tasks nicht konfiguriert ist, liest der Auftrags-Client nur den `tasksperslot`-Parameter aus. Sie können die Anzahl der Tasks jederzeit überschreiben – entweder über eine Bootstrap-Aktion für alle Cluster, oder einzeln pro Auftrag durch das Hinzufügen eines Schritts zum Ändern der Konfiguration.

Amazon EMR widersteht Ausfällen von Task-Knoten und setzt die Cluster-Ausführung fort, auch wenn ein Task-Knoten nicht mehr verfügbar ist. Amazon stellt EMR automatisch zusätzliche Task-Knoten bereit, um diejenigen zu ersetzen, die ausfallen.

Sie können für jeden Cluster-Schritt eine andere Anzahl von Aufgabenknoten nutzen. Sie können außerdem in einem ausgeführten Cluster einen Schritt hinzufügen, um die Anzahl der Aufgabenknoten zu ändern. Da alle Schritte standardmäßig sequenziell ausgeführt werden, können Sie die Anzahl der ausgeführten Aufgabenknoten für einen Schritt festlegen.

Installation von Anwendungen mit früheren AMI Versionen von Amazon EMR

[Wenn Sie eine AMI Version verwenden, werden Anwendungen auf verschiedene Arten installiert, z. B. mithilfe des `NewSupportedProducts` Parameters für die `RunJobFlow` Aktion, mithilfe von Bootstrap-Aktionen und mithilfe der Step-Aktion.](#)

Anpassen der Cluster- und Anwendungskonfiguration mit früheren AMI Versionen von Amazon EMR

Mit der EMR Amazon-Version 4.0.0 wurde eine vereinfachte Methode zur Konfiguration von Anwendungen mithilfe von Konfigurationsklassifizierungen eingeführt. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#). Wenn Sie eine AMI Version verwenden, konfigurieren Sie Anwendungen mithilfe von Bootstrap-Aktionen zusammen mit Argumenten, die Sie übergeben. Die Aktionen `configure-hadoop` und `configure-daemons` Bootstrap legen beispielsweise

Hadoop- und YARN spezifische Umgebungseigenschaften wie fest. `--namenode-heap-size` In neueren Versionen werden diese mit den Konfigurationsklassifizierungen `hadoop-env` und `yarn-env` konfiguriert. [Bootstrap-Aktionen, die allgemeine Konfigurationen durchführen, finden Sie im Repository auf Github. `emr-bootstrap-actions`](#)

In den folgenden Tabellen werden Bootstrap-Aktionen den Konfigurationsklassifizierungen in neueren EMR Amazon-Release-Versionen zugeordnet.

Hadoop

Dateiname der betroffenen Anwendung	AMIVersion: Bootstrap-Aktion	Konfigurationsklassifizierung
<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>
<code>hdfs-site.xml</code>	<code>configure-hadoop -s</code>	<code>hdfs-site</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

Dateiname der betroffenen Anwendung	AMIVersion: Bootstrap-Aktion	Konfigurationsklassifizierung
<code>hive-env.sh</code>	–	<code>hive-env</code>

Dateiname der betroffenen Anwendung	AMI Version: Bootstrap-Aktion	Konfigurationsklassifizierung
hive-site.xml	hive-script --install -hive-site \${MY_HIVE_SITE_FILE}	hive-site
hive-exec-log4j.properties	-	hive-exec-log4j
hive-log4j.properties	-	hive-log4j

EMRFS

Dateiname der betroffenen Anwendung	AMI Version: Bootstrap-Aktion	Konfigurationsklassifizierung
emrfs-site.xml	configure-hadoop -e	emrfs-site
-	s3get -s s3://cust-om-provider.jar -d /usr/share/aws/emr/auxlib/	emrfs-site (mit neuer Einstellung fs.s3.cse.encryptionMaterialsProvider.uri)

Eine Liste aller Klassifizierungen finden Sie unter [Anwendungen konfigurieren](#).

Variablen für die Anwendungsumgebung

Wenn Sie eine AMI Version verwenden, wird ein `hadoop-user-env.sh` Skript zusammen mit der `configure-daemons` Bootstrap-Aktion verwendet, um die Hadoop-Umgebung zu konfigurieren. Das Skript umfasst die folgenden Aktionen:

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```


In EMR Amazon-Version 4.x machen Sie dasselbe mit der `hadoop-env` Konfigurationsklassifizierung, wie im folgenden Beispiel gezeigt:

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
      },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_USER_CLASSPATH_FIRST": "true",
          "HADOOP_CLASSPATH": "/path/to/my.jar"
        }
      }
    ]
  }
]
```

Ein weiteres Beispiel verdeutlicht, dass die Verwendung von `configure-daemons` und die Übergabe von `--namenode-heap-size=2048` und `--namenode-opts=-XX:GCTimeRatio=19` den folgenden Konfigurationsklassifizierungen entspricht.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
      },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        }
      }
    ]
  }
]
```

Andere Umgebungsvariablen der Anwendung werden nicht länger in `/home/hadoop/.bashrc` definiert. Stattdessen werden sie hauptsächlich in `/etc/default`-Dateien nach Komponente oder Anwendung eingerichtet, z. B. `/etc/default/hadoop`. Wrapper-Skripte, die von der Anwendung `/usr/bin/` installiert werden, RPMs können auch zusätzliche Umgebungsvariablen setzen, bevor das eigentliche Bin-Skript verwendet wird.

Service-Ports

Wenn Sie eine AMI Version verwenden, verwenden einige Dienste benutzerdefinierte Ports.

Änderungen der Port-Einstellungen

Einstellung	AMIVersion 3.x	Open-Source-Standard
<code>fs.default.name</code>	<code>hdfs://IP: 9000 emrDetermined</code>	Standard (<code>hdfs://<i>emrDeterminedIP</i> :8020</code>)
<code>dfs.datanode.address</code>	<code>0.0.0.0:9200</code>	Standard (<code>0.0.0.0:50010</code>)
<code>dfs.datanode.http.address</code>	<code>0.0.0.0:9102</code>	Standard (<code>0.0.0.0:50075</code>)
<code>dfs.datanode.https.address</code>	<code>0.0.0.0:9402</code>	Standard (<code>0.0.0.0:50475</code>)
<code>dfs.datanode.ipc.address</code>	<code>0.0.0.0:9201</code>	Standard (<code>0.0.0.0:50020</code>)
<code>dfs.http.address</code>	<code>0.0.0.0:9101</code>	Standard (<code>0.0.0.0:50070</code>)
<code>dfs.https.address</code>	<code>0.0.0.0:9202</code>	Standard (<code>0.0.0.0:50470</code>)
<code>dfs.secondary.http.address</code>	<code>0.0.0.0:9104</code>	Standard (<code>0.0.0.0:50090</code>)
<code>yarn.nodemanager.address</code>	<code>0.0.0.0:9103</code>	Standard (<code>\${yarn.nodemanager.hostname}:0</code>)
<code>yarn.nodemanager.localizer.address</code>	<code>0.0.0.0:9033</code>	Standard (<code>\${yarn.nodemanager.hostname}:8040</code>)
<code>yarn.nodemanager.webapp.address</code>	<code>0.0.0.0:9035</code>	Standard (<code>\${yarn.nodemanager.hostname}:8042</code>)

Einstellung	AMI Version 3.x	Open-Source-Standard
yarn.resourcemanager.addresses	<i>emrDeterminedIP</i> :9022	Standard (\${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	<i>emrDeterminedIP</i> :9025	Standard (\${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	<i>emrDeterminedIP</i> :9023	Standard (\${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	<i>emrDeterminedIP</i> :9024	Standard (\${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	0.0.0.0:9026	Standard (\${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	<i>emrDeterminedIP</i> :9046	Standard (no-value)
yarn.resourcemanager.hostname	0.0.0.0 (Standard)	<i>emrDeterminedIP</i>

Note

Das Tool *emrDeterminedIP* ist eine IP-Adresse, die von Amazon generiert wird EMR.

Benutzer

Bei Verwendung einer AMI Version `hadoop` führt der Benutzer alle Prozesse aus und besitzt alle Dateien. In EMR Amazon-Release-Version 4.0.0 und höher existieren Benutzer auf Anwendungs- und Komponentenebene.

Installationssequenz, installierte Artefakte und Speicherorte für die Protokolldatei

Wenn Sie eine AMI Version verwenden, werden Anwendungsartefakte und ihre Konfigurationsverzeichnisse im `/home/hadoop/application` Verzeichnis installiert. Wenn Sie beispielsweise das Verzeichnis Hive installiert hatten, lautete das Verzeichnis `/home/hadoop/hive`. In EMR Amazon-Version 4.0.0 und höher werden Anwendungsartefakte im /

`usr/lib/application` Verzeichnis installiert. Wenn Sie eine AMI Version verwenden, befinden sich Protokolldateien an verschiedenen Stellen. In der folgenden Tabelle sind diese Speicherorte aufgeführt.

Änderungen der Protokollspeicherorte in Amazon S3

Daemon oder Anwendung	Speicherort des Verzeichnisses
instance-state	knoten/ <i>instance-id</i> /instanzzustand/
hadoop-hdfs-namenode	Dämonen/ <i>instance-id</i> /.log hadoop-hadoop-namenode
hadoop-hdfs-datanode	Dämonen/ <i>instance-id</i> /.log hadoop-hadoop-datanode
hadoop-Garn () ResourceManager	Dämonen/ <i>instance-id</i> /yarn-hadoop-resource-manager
hadoop-yarn (Proxy-Server)	Dämonen/ <i>instance-id</i> /yarn-hadoop-proxy-server
mapred-historyserver	Dämonen/ <i>instance-id</i> /
httpfs	Dämonen/ <i>instance-id</i> /httpfs.log
hive-server	knoten/ <i>instance-id</i> /hive-server/hive-server.log
hive-metastore	knoten/ <i>instance-id</i> /apps/hive.log
Bienenstock CLI	knoten/ <i>instance-id</i> /apps/hive.log
YARNAnwendungen, Benutzerprotokolle und Container-Protokolle	task-attempts/
Mahout	N/A
Pig	N/A
spark-historyserver	–

Daemon oder Anwendung	Speicherort des Verzeichnisses
Verlaufsdateien des MapReduce-Auftrags	jobs/

Command Runner

Wenn Sie eine AMI Version verwenden `/home/hadoop/contrib/streaming/hadoop-streaming.jar`, werden viele Skripte oder Programme nicht in der Shell-Anmeldepfadumgebung platziert. Sie müssen also den vollständigen Pfad angeben, wenn Sie eine JAR-Datei wie `command-runner.jar` oder `script-runner.jar` verwenden, um die Skripten auszuführen. Das `command-runner.jar` befindet sich auf der, AMI sodass Sie nicht alles wissen müssen, URI wie es bei der Fall `wascript-runner.jar`.

Replikationsfaktor

Mit dem Replikationsfaktor können Sie konfigurieren, wann ein Hadoop JVM gestartet werden soll. Sie können JVM für jede Aufgabe ein neues Hadoop starten, was für eine bessere Aufgabenisolierung sorgt, oder Sie können die Aufgaben gemeinsam nutzen, wodurch JVMs der Framework-Overhead geringer wird. Wenn Sie viele kleine Dateien verarbeiten, ist es sinnvoll, diese JVM mehrfach wiederzuverwenden, um die Startkosten zu amortisieren. Wenn jedoch jede Aufgabe viel Zeit in Anspruch nimmt oder eine große Datenmenge verarbeitet, sollten Sie sich dafür entscheiden, die nicht wiederzuverwenden, JVM um sicherzustellen, dass der gesamte Speicher für nachfolgende Aufgaben freigegeben wird. Wenn Sie eine AMI Version verwenden, können Sie den Replikationsfaktor mithilfe der `configure-hadoop Bootstrap`-Aktion anpassen, um die `mapred.job.reuse.jvm.num.tasks` Eigenschaft festzulegen.

Das folgende Beispiel zeigt, wie der JVM Wiederverwendungsfaktor für unendliche JVM Wiederverwendung festgelegt wird.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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"]
```

Hive-Anwendungsspezifikationen für frühere AMI Versionen von Amazon EMR

Protokolldateien

Bei Verwendung der EMR AMI Amazon-Versionen 2.x und 3.x werden Hive-Protokolle unter gespeichert. `/mnt/var/log/apps/` Zur Unterstützung gleichzeitiger Versionen von Hive bestimmt die Hive-Version, die Sie ausführen, den Protokolldateinamen, wie in der folgenden Tabelle dargestellt.

Hive-Version	Protokolldateiname
0.13.1	hive.log
	<div data-bbox="620 1144 657 1180"></div> Note Ab dieser Version EMR verwendet Amazon einen unversionierten Dateinamen, <code>hive.log</code> . Nebenversionen verwenden denselben Protokollspeicherort wie die Hauptversion.
0.11.0	hive_0110.log
	<div data-bbox="620 1575 657 1610"></div> Note Nebenversionen von Hive 0.11.0, z. B. 0.11.0.1, nutzen denselben Speicherort für ihre Protokolldatei wie Hive 0.11.0.
0.8.1	hive_081.log

Hive-Version	Protokolldateiname
	<p>Note</p> <p>Nebenversionen von Hive 0.8.1, z. B. Hive 0.8.1.1, nutzen denselben Speicherort für ihre Protokolldatei wie Hive 0.8.1.</p>
0.7.1	<p>hive_07_1.log</p> <p>Note</p> <p>Nebenversionen von Hive 0.7.1, z. B. Hive 0.7.1.3 und Hive 0.7.1.4, nutzen denselben Speicherort für ihre Protokolldatei wie Hive 0.7.1.</p>
0.7	hive_07.log
0.5	hive_05.log
0.4	hive.log

Split-Eingabe-Funktionen

Verwenden Sie Folgendes, um Split-Input-Funktionen mit Hive-Versionen vor 0.13.1 (EMRAMIAmazon-Versionen vor 3.11.0) zu implementieren:

```
hive> set hive.input.format=org.apache.hadoop.hive.q1.io.HiveCombineSplitsInputFormat;
hive> set mapred.min.split.size=100000000;
```

Diese Funktionalität wurde mit Hive 0.13.1 eingestellt. Verwenden Sie Folgendes, um dieselbe Funktionalität für das geteilte Eingabeformat in Amazon EMR AMI Version 3.11.0 zu erhalten:

```
set hive.hadoop.supports.plittable.combineinputformat=true;
```

Thrift-Service-Ports

Thrift ist ein RPC Framework, das ein kompaktes binäres Serialisierungsformat definiert, das verwendet wird, um Datenstrukturen für spätere Analysen beizubehalten. In der Regel konfiguriert Hive die Server für den Betrieb an folgenden Ports.

Hive-Version	Port-Nummer
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

Weitere Informationen über Thrift-Services finden Sie unter <http://wiki.apache.org/thrift/>.

Verwenden von Hive zum Wiederherstellen von Partitionen

Amazon EMR enthält eine Anweisung in der Hive-Abfragesprache, die die Partitionen einer Tabelle aus Tabellendaten in Amazon S3 wiederherstellt. Im folgenden Beispiel ist diese Anweisung dargestellt.

```
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;
```

Die Partitionsverzeichnisse und -daten müssen sich den dem in der Tabellendefinition angegebenen Speicherort befinden und dem Hive-Standard entsprechend benannt sein: z. B. dt=2009-01-01.

Note

Nach der Version Hive 0.13.1 wird diese Funktion nativ mit `msck repair table` unterstützt. Daher wird `recover partitions` nicht unterstützt. Weitere Informationen finden Sie unter <https://cwiki.apache.org/confluence/display/Hive/LanguageManual> +. DDL

Übergeben einer Hive-Variable an ein Skript

Um eine Variable mit dem an einen Hive-Schritt zu übergeben AWS CLI, geben Sie den folgenden Befehl ein: replace *myKey* mit dem Namen Ihres EC2 key pair und ersetzen *mybucket* mit Ihrem Bucket-Namen. In diesem Beispiel ist SAMPLE ein Variablenwert, dem der Switch `-d` vorangestellt wird. Diese Variable ist im Hive-Skript definiert als: `${SAMPLE}`.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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/]
```

Angeben eines externen Metastore-Speicherorts

Das folgende Verfahren zeigt, wie Sie die Standardkonfigurationswerte für den Speicherort des Hive-Metastores überschreiben und einen Cluster mit dem neu konfigurierten Metastore-Speicherort starten.

Um einen Metastore außerhalb des Clusters EMR zu erstellen

1. Erstellen Sie mit Amazon eine My SQL - oder Aurora-DatenbankRDS.

Informationen zum Erstellen einer RDS Amazon-Datenbank finden Sie unter [Erste Schritte mit Amazon RDS](#).

2. Ändern Sie Ihre Sicherheitsgruppen, um JDBC Verbindungen zwischen Ihrer Datenbank und der ElasticMapReduce-Master-Sicherheitsgruppe zuzulassen.

Informationen darüber, wie Sie Ihre Sicherheitsgruppen für den Zugriff ändern können, finden Sie unter [RDSAmazon-Sicherheitsgruppen](#) im RDSAmazon-Benutzerhandbuch.


3. Stellen Sie die JDBC Konfigurationswerte ein in `hive-site.xml`:

- a. Erstellen Sie eine `hive-site.xml`-Konfigurationsdatei, die folgende Informationen enthält:

```
<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 ist die DNS Adresse der RDS Amazon-Instance, auf der die Datenbank läuft. *username* und *password* sind die Anmeldeinformationen für Ihre Datenbank. Weitere Informationen zum Herstellen einer Verbindung zu My SQL - und Aurora-Datenbank-Instances finden Sie unter [Verbindung zu einer DB-Instance herstellen, auf der die My SQL Database-Engine ausgeführt wird](#), und [Herstellen einer Verbindung zu einem Aurora-DB-Cluster](#) im RDSAmazon-Benutzerhandbuch.


Die JDBC Treiber werden von Amazon installiert.

 Note

Die Werteeigenschaft darf keine Leerzeichen oder Wagenrückläufe enthalten. Sie sollte in einer Zeile erscheinen.

- b. Speichern Sie die `hive-site.xml`-Datei in einem Speicherort in Amazon S3, z. B. unter `s3://mybucket/hive-site.xml`.
4. Erstellen Sie einen Cluster, indem Sie den Amazon-S3-Speicherort der benutzerdefinierten Datei `hive-site.xml` angeben.

Der folgende Beispielbefehl demonstriert einen AWS CLI Befehl, der dies tut.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.10 \  
--applications Name=Hue Name=Hive Name=Pig \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m3.xlarge --instance-count 3 \  
--bootstrap-actions Name="Install Hive Site Configuration",\  
Path="s3://region.elasticmapreduce/libs/hive/hive-script",\  
Args=["--base-path", "s3://elasticmapreduce/libs/hive", "--install-hive-site",\  
"--hive-site=s3://mybucket/hive-site.xml", "--hive-versions", "latest"]
```

Connect zu Hive her mit JDBC

Um eine Verbindung zu Hive herzustellen, JDBC müssen Sie den JDBC Treiber herunterladen und einen SQL Client installieren. Das folgende Beispiel zeigt die Verwendung von SQL Workbench/J zum Herstellen einer Verbindung zu Hive mithilfe von JDBC

JDBC Um Treiber herunterzuladen

- Laden Sie die entsprechenden Treiber für die Hive-Versionen herunter, auf die Sie zugreifen möchten. Entpacken Sie die Treiber. Die Hive-Version unterscheidet sich je nachdem AMI, welche Sie bei der Erstellung eines EMR Amazon-Clusters auswählen.
 - [Hive JDBC 0.13.1-Treiber: _1.0.4.1004.zip](https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC) [https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/ AmazonHive JDBC](https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC)
 - Hive JDBC 0.11.0-Treiber: <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0>
 - Hive 0.8.1-Treiber: JDBC <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1>
- Installieren Sie SQL Workbench/J. Weitere Informationen finden Sie unter [Workbench/J installieren und starten im SQL Workbench/J](#) Manual User's Manual. SQL
- Erstellen Sie einen Tunnel zum Cluster-Masterknoten. SSH Der Port für die Verbindung unterscheidet sich je nach Hive-Version. In den folgenden Tabellen finden Sie Beispielbefehle für ssh Linux-Benutzer und TTY Pu-Befehle für Windows-Benutzer

SSH Linux-Befehle

Hive-Version	Befehl
0.13.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>
0.11.0	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10004:localhost:10004 hadoop@ <i>master-public-dns-name</i></code>
0.8.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10003:localhost:10003 hadoop@ <i>master-public-dns-name</i></code>
0.7.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10002:localhost:10002 hadoop@ <i>master-public-dns-name</i></code>
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>

Windows TTY PU-Tunneleinstellungen

Hive-Version	Tunneleinstellungen
0.13.1	Quellport: 10000 Ziel: <i>master-public-dns-name</i> : 10000
0.11.0	Quellport: 10004 Ziel: <i>master-public-dns-name</i> :10004
0.8.1	Quellport: 10003 Ziel: <i>master-public-dns-name</i> :10003

4. Fügen Sie den JDBC Treiber zu Workbench hinzu. SQL

- Klicken Sie im Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) auf Manage Drivers (Treiber verwalten).
- Wählen Sie das Symbol (leere Seite) Create a new entry (Neuen Eintrag erstellen) aus.
- Geben Sie in das Feld Name **Hive JDBC** ein.
- Klicken Sie unter Bibliothek auf das Symbol JARDatei (en) auswählen.
- Wählen Sie JAR Dateien aus, wie in der folgenden Tabelle dargestellt.

Hive-Treiberversion	JARhinzuzufügende Dateien
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</pre>

Hive-Treiberversion	JAR hinzuzufügende Dateien
	<pre>hive-jdbc-0.11.0.jar hive-metastore-0.11.0.jar hive-service-0.11.0.jar libfb303-0.9.0.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar</pre>
0.8.1	<pre>hadoop-core-0.20.205.jar hive-exec-0.8.1.jar hive-jdbc-0.8.1.jar hive-metastore-0.8.1.jar hive-service-0.8.1.jar libfb303-0.7.0.jar libthrift-0.7.0.jar log4j-1.2.15.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>
0.7.1	<pre>hadoop-0.20-core.jar hive-exec-0.7.1.jar hive-jdbc-0.7.1.jar hive-metastore-0.7.1.jar hive-service-0.7.1.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>

Hive-Treiberversion	JAR hinzuzufügende Dateien
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. Wählen Sie im Dialogfeld Please select one driver (Bitte einen Treiber auswählen) einen Treiber gemäß der folgenden Tabelle aus und klicken Sie dann auf OK.

Hive-Version	Treiber-Klassenname
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>

Hive-Version	Treiber-Klassenname
0.7	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.5	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

5. Wenn Sie zum Dialogfeld „Verbindungsprofil auswählen“ zurückkehren, stellen Sie sicher, dass das Feld Treiber auf Hive gesetzt ist, JDBC und geben Sie die JDBC Verbindungszeichenfolge gemäß der folgenden Tabelle in das URL-Feld ein.

Hive-Version	JDBC Verbindungszeichenfolge
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>

Wenn Ihr Cluster AMI Version 3.3.1 oder höher verwendet, geben Sie im Dialogfeld „Verbindungsprofil auswählen“ das **hadoop** Feld Benutzername ein.

HBase Anwendungsspezifikationen für frühere AMI Versionen von Amazon EMR

Unterstützte Versionen HBase

HBaseVersion	AMIVersion	AWS CLI Konfigurationsparameter	HBase Versionsdetails
0.94.18	3.1.0 und höher	<code>--ami-version 3.1</code>	<ul style="list-style-type: none"> Fehlerbehebungen und Verbesserungen

HBaseVersion	AMIVersion	AWS CLI Konfigurationsparameter	HBaseVersionsdetails
		<pre>--ami-version 3.2 --ami-version 3.3 --applications Name=HBase</pre>	
0.94.7	3.0-3.0.4	<pre>--ami-version 3.0 --applications Name=HBase</pre>	
0.92	2.2 und höher	<pre>--ami-version 2.2 or later --applications Name=HBase</pre>	

HBaseVoraussetzungen für den Cluster

Ein Cluster, der mit den EMR AMI Amazon-Versionen 2.x und 3.x erstellt wurde, sollte die folgenden Anforderungen erfüllen für HBase

- Das AWS CLI (optional) — Um mit der HBase Befehlszeile zu interagieren, laden Sie die neueste Version von [herunter](#) und installieren Sie sie. AWS CLI Weitere Informationen finden Sie unter [Installieren der AWS Command Line Interface](#) im AWS Command Line Interface - Benutzerhandbuch.
- Mindestens zwei Instanzen (optional) — Auf dem Master-Knoten des Clusters werden der HBase Master-Server und Zookeeper ausgeführt, und auf Task-Knoten werden die HBase Regionserver ausgeführt. Um eine optimale Leistung zu erzielen, sollten HBase Cluster auf mindestens zwei EC2 Instances ausgeführt werden. Zu Testzwecken können Sie sie jedoch auch HBase auf einem einzelnen Knoten ausführen.

- Cluster mit langer Laufzeit — läuft HBase nur auf Clustern mit langer Laufzeit. Standardmäßig erstellen die CLI und die EMR Amazon-Konsole lang andauernde Cluster.
- Ein EC2 Amazon-Schlüsselpaarsatz (empfohlen) — Um das Secure Shell (SSH) - Netzwerkprotokoll für die Verbindung mit dem Master-Knoten und die Ausführung von HBase Shell-Befehlen zu verwenden, müssen Sie bei der Erstellung des Clusters ein EC2 Amazon-Schlüsselpaar verwenden.
- Die richtige Version AMI und die Hadoop-Version — HBase Cluster werden derzeit nur auf Hadoop 20.205 oder höher unterstützt.
- Ganglia (optional) — Um HBase Leistungskennzahlen zu überwachen, installieren Sie Ganglia, wenn Sie den Cluster erstellen.
- Ein Amazon S3 S3-Bucket für Protokolle (optional) — Die Protokolle für HBase sind auf dem Master-Knoten verfügbar. Wenn Sie möchten, dass diese Protokolle nach Amazon S3 kopiert werden, geben Sie beim Erstellen des Clusters einen S3-Bucket für den Empfang der Protokolldateien an.

Erstellen eines Clusters mit HBase

In der folgenden Tabelle sind Optionen aufgeführt, die verfügbar sind, wenn Sie die Konsole verwenden, um einen Cluster mit HBase einer EMR AMI Amazon-Release-Version zu erstellen.

Feld	Aktion
Restore from backup (Wiederherstellen mithilfe der Sicherung)	Geben Sie an, ob der HBase Cluster mit in Amazon S3 gespeicherten Daten vorab geladen werden soll.
Backup location (Speicherort der Sicherung)	Geben Sie an URI, wo sich das Backup, von dem aus die Wiederherstellung erfolgen soll, in Amazon S3 befindet.
Backup version (Sicherungsversion)	Geben Sie optional den Versionsnamen der zu verwendenden Sicherung unter Backup location (Speicherort der Sicherung) an. Wenn Sie dieses Feld leer lassen, EMR verwendet Amazon das neueste Backup am Backup-Speicherort, um den neuen HBase Cluster zu füllen.

Feld	Aktion
Schedule regular backups (Regelmäßige Sicherungen planen)	Geben Sie an, ob automatisch inkrementelle Sicherungen eingeplant werden sollen. Die erste Sicherung ist ein vollständiges Backup, um eine Baseline für künftige inkrementelle Sicherungen zu erstellen.
Consistent backup (Konsistente Sicherung)	Geben Sie an, ob die Sicherungen konsistent sein sollen. Bei einer konsistenten Sicherung werden die Schreibvorgänge während der anfänglichen Sicherungsstufe, der knotenübergreifenden Synchronisierung, angehalten. Alle auf diese Weise angehaltenen Schreibvorgänge werden in eine Warteschlange gestellt und nach Abschluss der Synchronisierung fortgesetzt.
Backup frequency (Sicherungshäufigkeit)	Die Anzahl der Tage/Stunden/Minuten zwischen den geplanten Sicherungen.
Backup location (Speicherort der Sicherung)	Das Amazon S3URI, in dem Backups gespeichert werden. Der Backup-Speicherort für jeden HBase Cluster sollte unterschiedlich sein, um sicherzustellen, dass die differenziellen Backups korrekt bleiben.
Backup start time (Startzeit der Sicherung)	Geben Sie an, wann die erste Sicherung durchgeführt werden soll. Sie können dies auf <code>instantnow</code> einstellen, sodass das erste Backup gestartet wird, sobald der Cluster läuft, oder Sie können ein Datum und eine Uhrzeit im ISOFormat eingeben. Beispielsweise würde <code>2012-06-15T 20:00 Z</code> die Startzeit auf den 15. Juni 2012 um 20 Uhr festlegen. UTC

Mit dem folgenden AWS CLI Beispielbefehl wird ein Cluster mit und anderen Anwendungen gestartet:
HBase

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Nachdem die Verbindung zwischen Hive und HBase Clustern hergestellt wurde (wie im vorherigen Verfahren gezeigt), können Sie auf die im HBase Cluster gespeicherten Daten zugreifen, indem Sie eine externe Tabelle in Hive erstellen.

Wenn das folgende Beispiel über die Hive-Eingabeaufforderung ausgeführt wird, wird eine externe Tabelle erstellt, die auf Daten verweist, die in einer HBase Tabelle mit dem Namen gespeichert sind. `inputTable` Anschließend können Sie `inputTable` in Hive-Anweisungen referenzieren, um im Cluster gespeicherte Daten abzufragen und zu ändern. HBase

Note

Im folgenden Beispiel wird `protobuf-java-2.4.0a.jar` in AMI 2.3.3 verwendet, aber Sie sollten das Beispiel an Ihre Version anpassen. Um zu überprüfen, welche Version der Protocol Buffers JAR Sie haben, führen Sie den Befehl an der Hive-Befehlszeile aus: `! 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 ;
```

Konfiguration anpassen HBase

Obwohl die Standardeinstellungen für die meisten Anwendungen funktionieren sollten, haben Sie die Flexibilität, Ihre HBase Konfigurationseinstellungen zu ändern. Führen Sie hierzu eines der beiden Bootstrap-Aktionsskripts aus:

- `configure-hbase-daemons`— Konfiguriert die Eigenschaften der Master-, Regionalserver- und Zookeeper-Daemons. Zu diesen Eigenschaften gehören die Heap-Größe und Optionen, die beim Start des Daemons an die Java Virtual Machine () übergeben werden. JVM HBase Sie definieren diese Eigenschaften als Argumente in der Bootstrap-Aktion. Diese Bootstrap-Aktion ändert die Konfigurationsdatei `/home/hadoop/conf/.sh hbase-user-env` auf dem Cluster. HBase
- `configure-hbase` — Konfiguriert HBase standortspezifische Einstellungen wie den Port, an den der HBase Master eine Verbindung herstellen soll, und die maximale Häufigkeit, mit der der Client-Client eine Aktion wiederholen soll. CLI Sie können diese one-by-one als Argumente in der Bootstrap-Aktion festlegen oder den Speicherort einer XML Konfigurationsdatei in Amazon S3 angeben. Diese Bootstrap-Aktion ändert die Konfigurationsdatei `/home/hadoop/conf/hbase-site.xml` auf dem Cluster. HBase

Note

Diese Skripts können, wie andere Bootstrap-Aktionen, nur ausgeführt werden, wenn der Cluster erstellt wird. Sie können sie nicht verwenden, um die Konfiguration eines HBase Clusters zu ändern, der gerade läuft.

Wenn Sie die Aktionen `configure-hbase` oder `configure-hbase-daemonsbootstrap` ausführen, überschreiben die von Ihnen angegebenen Werte die Standardwerte. Alle Werte, die Sie nicht explizit angeben, erhalten die Standardwerte.

Die Konfiguration HBase mit diesen Bootstrap-Aktionen entspricht der Verwendung von Bootstrap-Aktionen in Amazon EMR zur Konfiguration von Hadoop-Einstellungen und Hadoop-Daemon-Eigenschaften. Der Unterschied besteht darin, dass es HBase keine Speicheroptionen pro Prozess gibt. Stattdessen werden Speicheroptionen mit dem `--daemon-opts` Argument festgelegt, wobei *daemon* wird durch den Namen des zu konfigurierenden Daemons ersetzt.

Konfigurieren Sie Daemons HBase

Amazon EMR bietet eine Bootstrap-Aktion `3://region.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`, mit der Sie die Konfiguration von HBase Daemons ändern können, wobei *region* ist die Region, in der Sie Ihren Cluster starten. HBase

Um HBase Daemons mit dem zu konfigurieren AWS CLI, fügen Sie beim Starten des Clusters die `configure-hbase-daemons` Bootstrap-Aktion hinzu, um einen oder mehrere Daemons zu konfigurieren. HBase Sie können die folgenden Eigenschaften festlegen:

Property (Eigenschaft)	Description (Beschreibung)
<code>hbase-master-opts</code>	Optionen, die steuern, wie der JVM Master-Daemon ausgeführt wird. Falls gesetzt, überschreiben diese die OPTS Standardvariablen <code>HBASE _ MASTER _</code> .
<code>regionserver-opts</code>	Optionen, die steuern, wie der Regionserver-Daemon JVM ausgeführt wird. Falls gesetzt, überschreiben diese die OPTS Standardvariablen <code>HBASE _ REGIONSERVER _</code> .
<code>zookeeper-opts</code>	Optionen, die steuern, wie der Zookeeper-Daemon JVM ausgeführt wird. Falls gesetzt, überschreiben diese die Standardvariablen <code>HBASE _ ZOOKEEPER _ OPTS</code> .

Weitere Informationen zu diesen Optionen finden Sie in der HBase Dokumentation unter [hbase-env.sh](#).

Eine Bootstrap-Aktion zum Konfigurieren von Werten für `zookeeper-opts` und `hbase-master-opts` wird im folgenden Beispiel dargestellt.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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"]
```

Konfigurieren Sie HBase die Site-Einstellungen

Amazon EMR bietet eine Bootstrap-Aktion `s3://elasticmapreduce/bootstrap-actions/configure-hbase`, mit der Sie die Konfiguration ändern können. HBase Sie können Konfigurationswerte one-by-one als Argumente in der Bootstrap-Aktion festlegen oder den Speicherort einer XML Konfigurationsdatei in Amazon S3 angeben. Das Festlegen von Konfigurationswerten one-by-one ist nützlich, wenn Sie nur einige Konfigurationseinstellungen festlegen müssen. Sie mithilfe einer XML Datei festzulegen, ist nützlich, wenn Sie viele Änderungen vornehmen müssen oder wenn Sie Ihre Konfigurationseinstellungen zur Wiederverwendung speichern möchten.

Note

Sie können dem Amazon S3 S3-Bucket-Namen ein Regionspräfix voranstellens `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase`, z. B. wo *region* ist die Region, in der Sie Ihren HBase Cluster starten.

Diese Bootstrap-Aktion ändert die `/home/hadoop/conf/hbase-site.xml` Konfigurationsdatei auf dem HBase Cluster. Die Bootstrap-Aktion kann nur ausgeführt werden, wenn der HBase Cluster gestartet wird.

Weitere Informationen zu den HBase Site-Einstellungen, die Sie konfigurieren können, finden Sie in der HBase Dokumentation unter [Standardkonfiguration](#).

Legen Sie die `configure-hbase` Bootstrap-Aktion fest, wenn Sie den HBase Cluster starten, und geben Sie die `hbase-site.xml` zu ändernden Werte an.

Um individuelle HBase Site-Einstellungen mit dem zu spezifizieren AWS CLI

- Um die `hbase.hregion.max.filesize` Einstellung zu ändern, geben Sie den folgenden Befehl ein und ersetzen Sie *myKey* mit dem Namen Ihres EC2 Amazon-Schlüsselpaars.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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"]
```

Um HBase Seiteneinstellungen mit einer XML Datei anzugeben, verwenden Sie AWS CLI

1. Erstellen Sie eine benutzerdefinierte Version von `hbase-site.xml`. Ihre benutzerdefinierte Datei muss gültig sein XML. Um die Wahrscheinlichkeit von Fehlern zu verringern, beginnen Sie mit der Standardkopie von `hbase-site.xml`, die sich auf dem EMR HBase Amazon-Masterknoten unter `beendet/home/hadoop/conf/hbase-site.xml` befindet, und bearbeiten Sie eine Kopie dieser Datei, anstatt eine Datei von Grund auf neu zu erstellen. Sie können Ihrer neuen Datei einen neuen Namen geben oder den Dateinamen `hbase-site.xml` belassen.
2. Laden Sie die benutzerdefinierte Datei `hbase-site.xml` in einen Amazon-S3-Bucket hoch. Es sollten die Berechtigungen so eingerichtet sein, dass das AWS Konto, das den Cluster startet, auf die Datei zugreifen kann. Wenn das AWS Konto, das den Cluster startet, auch den Amazon S3 S3-Bucket besitzt, hat es Zugriff.
3. Legen Sie die Bootstrap-Aktion `configure-hbase` fest, wenn Sie den HBase Cluster starten, und geben Sie den Speicherort Ihrer benutzerdefinierten Datei an. `hbase-site.xml` Im folgenden Beispiel werden die Werte für die HBase Site-Konfiguration auf die in der Datei angegebenen Werte festgelegt. `s3://mybucket/my-hbase-site.xml` Geben Sie den folgenden Befehl ein, replace `myKey` mit dem Namen Ihres EC2 key pair und ersetzen `mybucket` mit dem Namen Ihres Amazon S3 S3-Buckets.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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://mybucket/my-hbase-site.xml,Args=["-s", "hbase.hregion.max.filesize=52428800"]
```



```
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type c1.xlarge --instance-count 3 --termination-protected \  
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["--site-config-file", "s3://mybucket/config.xml"]
```

Wenn Sie mehr als eine Option angeben, um den HBase Vorgang anzupassen, müssen Sie jedem Schlüssel-Wert-Paar einen `-s` Optionsschalter voranstellen, wie im folgenden Beispiel gezeigt:

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["-s", "zookeeper.session.timeout=60000"]
```

Wenn der Proxy eingerichtet und die SSH Verbindung geöffnet ist, können Sie die HBase Benutzeroberfläche anzeigen, indem Sie ein Browserfenster mit `http://öffnenmaster-public-dns-name:60010/master-status`, wo *master-public-dns-name* ist die öffentliche DNS Adresse des Master-Knotens im Cluster. HBase

Sie können die aktuellen HBase Protokolle anzeigenSSH, indem Sie eine Verbindung zum Master-Knoten herstellen und zum `mnt/var/log/hbase` Verzeichnis navigieren. Diese Protokolle sind nicht verfügbar, nachdem der Cluster beendet wurde, es sei denn, Sie aktivieren beim Starten des Clusters die Protokollierung nach Amazon S3.

Sichern und wiederherstellen HBase

Amazon EMR bietet die Möglichkeit, Ihre HBase Daten entweder manuell oder nach einem automatisierten Zeitplan auf Amazon S3 zu sichern. Sie können sowohl vollständige als auch inkrementelle Sicherungen durchführen. Nachdem Sie eine gesicherte Version der HBase Daten erstellt haben, können Sie diese Version in einem HBase Cluster wiederherstellen. Sie können die Wiederherstellung auf einem HBase Cluster durchführen, der gerade ausgeführt wird, oder einen neuen Cluster starten, der bereits mit gesicherten Daten gefüllt ist.

Führt während des Backup-Vorgangs HBase weiterhin Schreibbefehle aus. Dadurch wird zwar sichergestellt, dass Ihr Cluster während der Sicherung verfügbar bleibt, es besteht aber das Risiko von Inkonsistenzen zwischen den Daten, die gesichert werden, und den Schreibvorgängen, die parallel ausgeführt werden. Um die möglicherweise auftretenden Inkonsistenzen zu verstehen, müssen Sie berücksichtigen, dass Schreibvorgänge auf die Knoten im Cluster HBase verteilt werden. Wenn ein Schreibvorgang erfolgt, nachdem ein bestimmter Knoten abgefragt wurde, werden

diese Daten nicht in das Sicherungsarchiv aufgenommen. Möglicherweise stellen Sie sogar fest, dass frühere Schreibvorgänge in den HBase Cluster (die an einen Knoten gesendet wurden, der bereits abgefragt wurde) möglicherweise nicht im Backup-Archiv enthalten sind, während spätere Schreibvorgänge (die an einen Knoten gesendet wurden, bevor dieser abgerufen wurde) enthalten sind.

Wenn ein konsistentes Backup erforderlich ist, müssen Sie die Schreibvorgänge HBase während des ersten Teils des Backup-Vorgangs, d. h. der knotenübergreifenden Synchronisation, unterbrechen. Sie tun dies, indem Sie beim Anfordern einer Sicherung den Parameter `--consistent` angeben. Mit diesem Parameter werden die Schreibvorgänge während dieses Zeitraums in eine Warteschlange gestellt und ausgeführt, sobald die Synchronisierung abgeschlossen ist. Sie können auch regelmäßige Sicherungen einplanen. Dadurch werden Inkonsistenzen, die im Laufe der Zeit entstehen, aufgelöst, da die Daten, die bei einer Sicherung nicht berücksichtigt wurden, im folgenden Vorgang gesichert werden.

Wenn Sie HBase Daten sichern, sollten Sie für jeden Cluster ein anderes Backup-Verzeichnis angeben. Eine einfache Möglichkeit dafür ist die Verwendung der Cluster-Kennung als Bestandteil des für das Sicherungsverzeichnis angegebenen Pfads. Beispiel, `s3://mybucket/backups/j-3AEXXXXXX16F2`. Dadurch wird sichergestellt, dass alle future inkrementellen Backups auf den richtigen HBase Cluster verweisen.

Wenn Sie bereit sind, alte Sicherungsdateien zu löschen, die nicht mehr benötigt werden, empfehlen wir Ihnen, zunächst eine vollständige Sicherung Ihrer HBase Daten zu erstellen. Auf diese Weise wird sichergestellt, dass alle Daten erhalten bleiben. Außerdem bietet Ihnen dies eine Baseline für zukünftige inkrementelle Sicherungen. Nachdem die vollständige Sicherung abgeschlossen ist, können Sie zum Sicherungsverzeichnis navigieren und die alten Sicherungsdateien manuell löschen.

Der HBase Sicherungsvorgang verwendet S3 DistCp für den Kopiervorgang, der bestimmte Einschränkungen in Bezug auf den Speicherplatz für temporäre Dateien aufweist.

Sicherung und Wiederherstellung HBase mithilfe der Konsole

Die Konsole bietet die Möglichkeit, einen neuen Cluster zu starten und ihn mit Daten aus einem früheren HBase Backup zu füllen. Sie bietet Ihnen auch die Möglichkeit, regelmäßige inkrementelle Backups von HBase Daten zu planen. Zusätzliche Sicherungs- und Wiederherstellungsfunktionen, wie z. B. die Möglichkeit, Daten auf einem bereits laufenden Cluster wiederherzustellen, manuelle Backups durchzuführen und automatische vollständige Backups zu planen, sind mit dem CLI verfügbar.

So füllen Sie mithilfe der Konsole einen neuen Cluster mit archivierten HBase Daten

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Cluster erstellen.
3. Wählen Sie im Abschnitt Softwarekonfiguration für Zusätzliche Anwendungen die Option Konfigurieren HBase und hinzufügen aus.
4. Klicken Sie auf das Dialogfeld Add Application (Anwendung hinzufügen) und aktivieren Sie das Kontrollkästchen Restore From Backup (Wiederherstellen mithilfe der Sicherung).
5. Geben Sie Backup Backup-Speicherort den Speicherort des Backups an, das in den neuen HBase Cluster geladen werden soll. Dies sollte ein Amazon S3 URL der Form `s3://myawsbucket/backups/`.
6. Für Backup Version (Sicherungsversion) haben Sie durch Angabe eines Werts die Möglichkeit, den Namen einer zu ladenden Sicherungsversion anzugeben. Wenn Sie keinen Wert für Backup-Version festlegen, EMR lädt Amazon das neueste Backup an den angegebenen Speicherort.
7. Wählen Sie Add (Hinzufügen) und fahren Sie fort, um den Cluster mit den anderen Optionen wie gewünscht zu erstellen.

Um automatische HBase Datensicherungen mithilfe der Konsole zu planen

1. Wählen Sie im Abschnitt Softwarekonfiguration für Zusätzliche Anwendungen die Option Konfigurieren HBase und hinzufügen aus.
2. Wählen Sie Schedule Regular Backups (Regelmäßige Sicherungen planen).
3. Geben Sie an, ob die Sicherungen konsistent sein sollen. Bei einer konsistenten Sicherung werden die Schreibvorgänge während der anfänglichen Sicherungsstufe, der knotenübergreifenden Synchronisierung, angehalten. Alle auf diese Weise angehaltenen Schreibvorgänge werden in eine Warteschlange gestellt und nach Abschluss der Synchronisierung fortgesetzt.
4. Legen Sie fest, wie oft Sicherungen durchgeführt werden sollen, indem Sie eine Zahl für Backup Frequency (Sicherungshäufigkeit) eingeben, und wählen Sie Days (Tage), Hours (Stunden) oder Minutes (Minuten) aus. Das erste automatische Backup, das ausgeführt wird, ist ein vollständiges Backup. Danach EMR speichert Amazon inkrementelle Backups auf der Grundlage des von Ihnen angegebenen Zeitplans.

5. Geben Sie den Speicherort in Amazon S3 an, in dem die Sicherungen gespeichert werden sollen. Jeder HBase Cluster sollte an einem separaten Speicherort in Amazon S3 gesichert werden, um sicherzustellen, dass inkrementelle Backups korrekt berechnet werden.
6. Geben Sie an, wann die erste Sicherung erfolgen soll, indem Sie einen Wert für Backup Start Time (Startzeit der Sicherung) eingeben. Sie können dies auf `einstellennow`, wodurch das erste Backup gestartet wird, sobald der Cluster läuft, oder ein Datum und eine Uhrzeit im [ISOFormat](#) eingeben. Beispiel: `2013-09-26T 20:00 Z` legt die Startzeit auf den 26. September 2013 um 20 Uhr fest. UTC
7. Wählen Sie Hinzufügen aus.
8. Fahren Sie mit der Erstellung des Clusters je nach Bedarf mit den anderen Optionen fort.

Monitor mit HBase CloudWatch

Amazon EMR meldet drei Kennzahlen, mit CloudWatch denen Sie Ihre HBase Backups überwachen können. Diese Kennzahlen werden CloudWatch in Intervallen von fünf Minuten aktualisiert und kostenlos zur Verfügung gestellt.

Metrik	Beschreibung
HBaseBackupFailed	<p>Gibt an, ob die letzte Sicherung fehlgeschlagen ist. Sie ist standardmäßig auf 0 gesetzt und wird in 1 geändert, wenn der vorhergehende Sicherungsversuch fehlgeschlagen ist. Diese Metrik wird nur für HBase Cluster gemeldet.</p> <p>Anwendungsfall: HBase Backups überwachen</p> <p>Einheiten: Anzahl</p>
HBaseMostRecentBackupDuration	<p>Dauer, bis die vorherige Sicherung abgeschlossen ist. Diese Metrik wird unabhängig davon erstellt, ob die letzte Sicherung erfolgreich war oder fehlgeschlagen ist. Während die Sicherung ausgeführt wird, gibt die Metrik die Anzahl der Minuten seit dem Start der Sicherung zurück. Diese Metrik wird nur für HBase Cluster gemeldet.</p>

Metrik	Beschreibung
	Anwendungsfall: HBase Backups überwachen Einheiten: Minuten
HBaseTimeSinceLastSuccessfulBackup	Die Anzahl der Minuten, die nach dem Start des letzten erfolgreichen HBase Backups auf Ihrem Cluster vergangen sind. Diese Metrik wird nur für HBase Cluster gemeldet. Anwendungsfall: HBase Backups überwachen Einheiten: Minuten

Konfigurieren Sie Ganglia für HBase

Sie konfigurieren Ganglia für die HBase Verwendung der `configure-hbase-for-gangliaBootstrap`-Aktion. Diese Bootstrap-Aktion ist so konfiguriert, dass Metriken HBase in Ganglia veröffentlicht werden.

Sie müssen Ganglia konfigurieren HBase, wenn Sie den Cluster starten. Ganglia Reporting kann nicht zu einem laufenden Cluster hinzugefügt werden.

Ganglia speichert Protokolldateien auch auf dem Server unter `/mnt/var/log/ganglia/rrds`. Wenn Sie Ihren Cluster so konfiguriert haben, dass Protokolldateien dauerhaft in einem Amazon-S3-Bucket gespeichert werden, werden die Ganglia-Protokolldateien auch dort dauerhaft abgelegt.

Um einen Cluster mit Ganglia for zu starten HBase, verwenden Sie die `configure-hbase-for-gangliaBootstrap`-Aktion, wie im folgenden Beispiel gezeigt.

Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Nachdem der Cluster gestartet wurde, für den Ganglia konfiguriert ist, können Sie auf die Ganglia-Grafiken und -Berichte über die Benutzeroberfläche auf dem Master-Knoten zugreifen.

Pig-Anwendungsspezifikationen für frühere AMI Versionen von Amazon EMR

Unterstützte Pig-Versionen

Die Pig-Version, die Sie Ihrem Cluster hinzufügen können, hängt von der Version von Amazon EMR AMI und der Version von Hadoop ab, die Sie verwenden. Die folgende Tabelle zeigt, welche AMI Versionen und Versionen von Hadoop mit den verschiedenen Versionen von Pig kompatibel sind. Wir empfehlen die Verwendung der neuesten Pig-Version, um die Vorteile der Leistungsverbesserungen und neuen Funktionen zu nutzen.

Wenn Sie Pig API zur Installation von verwenden, wird die Standardversion verwendet, sofern Sie nicht `--pig-versions` als Argument für den Schritt angeben, der Pig während des Aufrufs von in den Cluster lädt. [RunJobFlow](#)

Pig-Version	AMIVersion	Konfigurationsparameter	Details zur Pig-Version
0.12.0 Versionshinweise Dokumentation	3.1.0 und höher	<code>--ami-version</code> 3.1 <code>--ami-version</code> 3.2 <code>--ami-version</code> 3.3	Implementiert die Unterstützung der folgenden Elemente: <ul style="list-style-type: none"> • Streaming UDFs ohne JVM Implementierungen • ASSERT und IN-Operatoren • CASE Ausdruck

Pig-Version	AMIVersion	Konfigurationsparameter	Details zur Pig-Version
			<ul style="list-style-type: none"> • AvroStorage als integrierte Pig-Funktion. • ParquetLoader und ParquetStorer als eingebaute Funktionen • BigInteger und BigDecimal Typen
0.11.1.1 Versionshinweise Dokumentation	2.2 und höher	<pre>--pig-versions 0.11.1.1 --ami-version 2.2</pre>	Verbessert die LOAD Befehlsleistung, PigStorage wenn sich die Eingabe in Amazon S3 befindet.
0.11.1 Versionshinweise Dokumentation	2.2 und höher	<pre>--pig-versions 0.11.1 --ami-version 2.2</pre>	Integriert die Unterstützung für JDK 7, Hadoop 2, benutzerdefinierte Groovy-Funktionen, SchemaTuple Optimierung, neue Operatoren und mehr. Weitere Informationen finden Sie im Veränderungsprotokoll für Pig 0.11.1 .
0.9.2.2 Versionshinweise Dokumentation	2.2 und höher	<pre>--pig-versions 0.9.2.2 --ami-version 2.2</pre>	Implementiert die Unterstützung von Hadoop 1.0.3.

Pig-Version	AMI-Version	Konfigurationsparameter	Details zur Pig-Version
0.9.2.1 Versionshinweise Dokumentation	2.2 und höher	<code>--pig-versions 0.9.2.1</code> <code>--ami-version 2.2</code>	Implementiert die Unterstützung von MapR.
0.9.2 Versionshinweise Dokumentation	2.2 und höher	<code>--pig-versions 0.9.2</code> <code>--ami-version 2.2</code>	Umfasst mehrere Leistungsverbesserungen und Fehlerkorrekturen. Die vollständigen Informationen zu den Änderungen mit Pig 0.9.2 finden Sie im Veränderungsprotokoll für Pig 0.9.2 .
0.9.1 Versionshinweise Dokumentation	2.0	<code>--pig-versions 0.9.1</code> <code>--ami-version 2.0</code>	
0.6 Versionshinweise	1,0	<code>--pig-versions 0.6</code> <code>--ami-version 1.0</code>	
0.3 Versionshinweise	1,0	<code>--pig-versions 0.3</code> <code>--ami-version 1.0</code>	

Details zur Pig-Version

Amazon EMR unterstützt bestimmte Pig-Versionen, für die möglicherweise zusätzliche EMR Amazon-Patches angewendet wurden. Sie können konfigurieren, welche Version von Pig auf EMR Amazon-Clustern ausgeführt werden soll. Weitere Information dazu finden Sie unter [Apache Pig](#). In den folgenden Abschnitten werden verschiedene Pig-Versionen und die Patches beschrieben, die auf die auf Amazon geladenen Versionen angewendet wurden EMR.

Pig-Patches

In diesem Abschnitt werden die benutzerdefinierten Patches beschrieben, die auf die bei Amazon verfügbaren Pig-Versionen angewendet wurden EMR.

Pig-0.11.1.1-Patches

Die EMR Amazon-Version von Pig 0.11.1.1 ist eine Wartungsversion, die die LOAD Befehlsleistung verbessert, PigStorage wenn sich die Eingabe in Amazon S3 befindet.

Pig 0.11.1-Patches

Die EMR Amazon-Version von Pig 0.11.1 enthält alle von der Apache Software Foundation bereitgestellten Updates und die kumulativen EMR Amazon-Patches von Pig Version 0.9.2.2. In Pig 0.11.1 gibt es jedoch keine neuen EMR Amazon-spezifischen Patches.

Pig-0.9.2-Patches

Apache Pig 0.9.2 ist eine Wartungsversion von Pig. Das EMR Amazon-Team hat die folgenden Patches auf die EMR Amazon-Version von Pig 0.9.2 angewendet.

Patch	Beschreibung
PIG-1429	<p>Fügt den Datentyp Boolean als Pig-Datentyp als primären Datentyp zu Pig hinzu. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/PIG-1429.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.10</p>

Patch	Beschreibung
PIG-1824	<p>Support Importmodule in JythonUDF. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -1824.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.10</p>
PIG-2010	<p>Das Paket ist JARs im verteilten Cache registriert. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -2010.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.11</p>
PIG-2456	<p>~/pigbootup-Datei hinzugefügt, über die Benutzer Standard-Pig-Anweisungen angeben können. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -2456.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.11</p>
PIG-2623	<p>Support der Verwendung von Amazon S3 S3-Pfaden zur RegistrierungUDFs. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -2623.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.10, 0.11</p>

Pig-0.9.1-Patches

Das EMR Amazon-Team hat die folgenden Patches auf die EMR Amazon-Version von Pig 0.9.1 angewendet.

Patch	Beschreibung
JARUnterstützungsdateien und Pig-Skripte in DFS	<p>Fügen Sie Unterstützung für das Ausführen von Skripten und das Registrieren von JAR Dateien hinzuHDFS, die in Amazon S3 oder anderen verteilten Dateisystemen gespeichert sind. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -1505.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.8.0</p>
Unterstützung mehrerer Dateisysteme in Pig	<p>Fügt Unterstützung für Pig-Skripts zum Lesen von Daten aus einem Dateisystem und Schreiben in einem anderen hinzu. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -1564.</p> <p>Status: Nicht ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) –</p>
Fügen Sie Piggybank Datetime und Zeichenfolge hinzu UDFs	<p>Fügen Sie Datetime und Zeichenfolge hinzu, um benutzerdefinierte Pig-Skripte UDFs zu unterstützen. Weitere Informationen finden Sie unter https://issues.apache.org/jira/browse/ PIG -1565.</p> <p>Status: Nicht ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) –</p>

Interaktive Pig-Cluster und Batch-Pig-Cluster

Amazon EMR ermöglicht es Ihnen, Pig-Skripte in zwei Modi auszuführen:

- Interactive
- Stapel

Wenn Sie einen Cluster mit langer Laufzeit über die Konsole oder die starten AWS CLI, können Sie eine Verbindung ssh zum Master-Knoten als Hadoop-Benutzer herstellen und die Grunt-Shell verwenden, um Ihre Pig-Skripte interaktiv zu entwickeln und auszuführen. Durch die interaktive Verwendung von Pig können Sie Pig-Skripts einfacher überarbeiten als im Batch-Modus. Nachdem Sie das Pig-Skript im interaktiven Modus erfolgreich überarbeitet haben, können Sie das Skript in Amazon S3 hochladen und den Batch-Modus nutzen, um das Skript in der Produktionsumgebung auszuführen. Sie können außerdem interaktiv Pig-Befehle an einen aktuell ausgeführten Cluster senden, um Daten bei Bedarf zu analysieren und zu transformieren.

Im Batch-Modus laden Sie in einem einzigen Schritt Ihr Pig-Skript in Amazon S3 hoch und senden gleichzeitig den Auftrag an den Cluster. Pig-Schritte können an lang- oder kurzlebige Cluster und an Übergangs-Cluster gesendet werden.

Spark-Anwendungsspezifikationen mit früheren AMI Versionen von Amazon EMR

Spark im interaktiven Modus oder im Batch-Modus verwenden

Amazon EMR ermöglicht es Ihnen, Spark-Anwendungen in zwei Modi auszuführen:

- Interactive
- Stapel

Wenn Sie einen Cluster mit langer Laufzeit über die Konsole oder die starten AWS CLI, können Sie SSH sich als Hadoop-Benutzer mit dem Master-Knoten verbinden und die Spark-Shell verwenden, um Ihre Spark-Anwendungen interaktiv zu entwickeln und auszuführen. Mit der interaktiven Verwendung von Spark ist das Entwickeln von Prototypen für Spark-Anwendungen und das Testen dieser Anwendungen einfacher als in einer Batch-Umgebung. Nachdem Sie die Spark-Anwendung erfolgreich im interaktiven Modus überarbeitet haben, können Sie diese Anwendung JAR oder das

Python-Programm in das lokale Dateisystem des Master-Knotens des Clusters auf Amazon S3 einfügen. Anschließend können Sie die Anwendung als Batch-Workflow senden.

Laden Sie im Batch-Modus das Spark-Skript in Amazon S3 oder ins Dateisystem des lokalen Hauptknotens hoch und senden Sie dann die Arbeit als Schritt an den Cluster. Spark-Schritte können an lang- oder kurzlebige Cluster gesendet werden.

Erstellen eines Clusters mit installiertem Spark

So starten Sie einen Cluster mit installiertem Spark über die Console

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option [Zur alten Konsole wechseln](#) aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Cluster erstellen.
3. Wählen Sie unter Softwarekonfiguration die AMI Release-Version aus, die Sie benötigen.
4. Wählen Sie unter Applications to be installed (Zu installierende Anwendungen) in der Liste Spark und anschließend Configure and add (Konfigurieren und hinzufügen) aus.
5. Sie können Argumente hinzufügen, um die Spark-Konfiguration bei Bedarf zu ändern. Weitere Informationen finden Sie unter [Konfigurieren von Spark](#). Wählen Sie Hinzufügen aus.
6. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Das folgende Beispiel zeigt, wie Sie einen Cluster mit Spark mittels Java erstellen:

```
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);
```

Konfigurieren von Spark

Sie konfigurieren Spark, wenn Sie einen Cluster erstellen, indem Sie die Bootstrap-Aktion ausführen, die sich im [emr-bootstrap-actionsaws/slabs/spark-Repository](#) auf Github befindet. Argumente, die von der Bootstrap-Aktion akzeptiert werden, finden Sie in diesem Repository. [README](#) Die Bootstrap-Aktion konfiguriert Eigenschaften in der Datei `$SPARK_CONF_DIR/spark-defaults.conf`. Weitere Informationen zu den Einstellungen finden Sie im Thema zur Spark-Konfiguration in der Spark-Dokumentation. Sie können „latest“ im Folgenden URL durch die Versionsnummer von Spark ersetzen, den Sie installieren, zum Beispiel. 2.2.0 <http://spark.apache.org/docs/latest/configuration.html>

Ebenso ist es möglich, Spark dynamisch bei jeder Anwendungsübermittlung zu konfigurieren. Eine Einstellung, mit der die Ressourcenzuordnung für einen Executor automatisch erhöht wird, ist über die Konfigurationsdatei `spark` verfügbar. Weitere Informationen finden Sie unter [Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen](#).

Ändern von Spark-Standardinstellungen

Das folgende Beispiel zeigt, wie Sie einen Cluster mit der AWS CLI erstellen, bei dem `spark.executor.memory` auf 2G gesetzt ist.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Aufträge an Spark übergeben

Um Arbeit an einen Cluster zu senden, führen Sie einen Schritt aus, um das `spark-submit` Skript auf Ihrem EMR Cluster auszuführen. Fügen Sie den Schritt mithilfe der `addJobFlowSteps` folgenden Methode hinzu [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);
```

Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen

Vielleicht möchten Sie die standardmäßigen Spark-Konfigurationswerte je nach Anwendung überschreiben. Sie können dies tun, wenn Sie Anwendungen mithilfe eines Schritts senden, wodurch im Wesentlichen Optionen an `spark-submit` übergeben werden. Sie möchten z. B. den Speicher ändern, der einem Executor-Prozess zugeordnet wurde. Dazu modifizieren Sie `spark.executor.memory`. Sie können den Schalter `--executor-memory` mit einem Argument wie dem folgenden angeben:

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

Auf ähnliche Weise können Sie `--executor-cores` und `--driver-memory` optimieren. In einem Schritt würden Sie die folgenden Argumente bereitstellen:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

Sie können mithilfe der Option `--conf` auch Einstellungen optimieren, die eventuell über keinen integrierten Schalter verfügen. Weitere Informationen zu anderen optimierbaren Einstellungen finden Sie im Thema [Dynamisches Laden von Spark-Eigenschaften](#) in der Apache-Spark-Dokumentation.

Unterschiede zwischen dem DistCp S3-Hilfsprogramm und früheren AMI Versionen von Amazon EMR

In Amazon unterstützte DistCp S3-Versionen EMR

Die folgenden DistCp S3-Versionen werden in EMR AMI Amazon-Versionen unterstützt. DistCpS3-Versionen nach 1.0.7 befinden sich direkt auf den Clustern. Verwenden Sie JAR den Link `/home/hadoop/lib` für die neuesten Funktionen.

Version	Beschreibung	Datum der Veröffentlichung
1.0.8	Fügt die Optionen <code>--appendToLastFile</code> , <code>--requirePreviousManifest</code> und <code>--storageClass</code> hinzu.	3. Januar 2014
1.0.7	Fügt die Option <code>--s3ServerSideEncryption</code> hinzu.	2. Mai 2013
1.0.6	Fügt die Option <code>--s3Endpoint</code> hinzu.	6. August 2012
1.0.5	Verbessert die Fähigkeit, anzugeben, welche Version von S3 ausgeführt werden DistCp soll.	27. Juni 2012
1.0.4	Verbessert die Option <code>--deleteOnSuccess</code> .	19. Juni 2012
1.0.3	Bietet Support für die Optionen <code>--numberOfFiles</code> und <code>--startingIndex</code> .	12. Juni 2012
1.0.2	Verbessert die Dateibenennung bei der Verwendung von Gruppen.	6. Juni 2012
1.0.1	Erste Version von S3DistCp.	19. Januar 2012

Fügen Sie einem Cluster einen DistCp S3-Kopierschritt hinzu

Um einem laufenden Cluster einen DistCp S3-Kopierschritt hinzuzufügen, geben Sie den folgenden Befehl ein: replace *j-3GYXXXXXX9I0K* mit Ihrer Cluster-ID und ersetzen *mybucket* mit Ihrem Amazon S3 S3-Bucket-Namen.

Note


Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
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,]+" ]
```

Example Laden Sie CloudFront Amazon-Logs in HDFS

In diesem Beispiel werden CloudFront Amazon-Logs geladen, HDFS indem ein Schritt zu einem laufenden Cluster hinzugefügt wird. Dabei wird das Komprimierungsformat von Gzip (CloudFrontStandard) auf LZO geändert. Dies ist nützlich, da Daten, die mit komprimiert wurden, beim Dekomprimieren in mehrere Maps aufgeteilt werden LZO können, sodass Sie nicht warten müssen, bis die Komprimierung abgeschlossen ist, wie dies bei Gzip der Fall ist. Dies bietet eine bessere Leistung, wenn Sie die Daten mit Amazon analysierenEMR. Dieses Beispiel verbessert außerdem die Leistung, indem der in der Option --groupBy angegebene reguläre Ausdruck verwendet wird, um alle Protokolle für eine bestimmte Stunde in einer einzigen Datei zu kombinieren. EMRAmazon-Cluster sind effizienter bei der Verarbeitung einiger großer LZO komprimierter Dateien als bei der Verarbeitung vieler kleiner, GZIP-komprimierter Dateien. Um LZO Dateien aufzuteilen, müssen Sie sie indizieren und die Hadoop-lzo-Bibliothek eines Drittanbieters verwenden.

Um CloudFront Amazon-Logs in zu ladenHDFS, geben Sie den folgenden Befehl ein: replace *j-3GYXXXXXX9I0K* mit Ihrer Cluster-ID und ersetzen *mybucket* mit Ihrem Amazon S3 S3-Bucket-Namen.

 Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
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"]
```

Stellen Sie sich den Fall vor, in dem das vorherige Beispiel über die folgenden CloudFront Protokolldateien ausgeführt wird.

```
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.HLUS3JKx.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.I9CNAZrg.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.YRRwERSA.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.dshVLXFE.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.LpLfuShd.gz
```

S3 DistCp kopiert, verkettet und komprimiert die Dateien in die folgenden beiden Dateien, wobei der Dateiname durch die Übereinstimmung mit dem regulären Ausdruck bestimmt wird.

```
hdfs:///local/2012-02-23-01.lzo
hdfs:///local/2012-02-23-02.lzo
```

Neuerungen?

Auf dieser Seite werden die Änderungen und Funktionen beschrieben, die in den neuesten Versionen von Amazon EMR 7.x, 6.x und 5.x verfügbar sind.

Diese Versionshinweise sind auch auf den Seiten [Amazon EMR 7.2.0](#), [Amazon EMR 6.15.0](#) und [Amazon EMR 5.36.2](#) verfügbar, zusammen mit den Anwendungsversionen, Komponentenversionen und verfügbaren Konfigurationsklassifizierungen für jede Version.

- Versionshinweise zu früheren Versionen finden Sie unter [EMR Amazon-Archiv mit Versionshinweisen](#).
- Um Updates zu erhalten, wenn eine neue EMR Amazon-Version verfügbar ist, abonnieren Sie den [RSSFeed für EMR Amazon-Versionshinweise](#).

Note

Spätere Versionen von Amazon EMR verwenden AWS Signature Version 4 (SigV4), um Anfragen an Amazon S3 zu authentifizieren. Wir empfehlen Ihnen, eine EMR Amazon-Version zu verwenden, die SigV4 dies unterstützt, damit Sie auf neue S3-Buckets zugreifen und Unterbrechungen Ihrer Workloads vermeiden können. Weitere Informationen und eine Liste der EMR Amazon-Versionen, die unterstützt werden SigV4, finden Sie unter [Amazon EMR und AWS Signature Version 4](#).

Amazon EMR 7.2.0 (neueste Version der 7.x-Serie)

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 7.2.0. Die Änderungen beziehen sich auf 7.2.0.

Neue Features

- Anwendungs-Upgrades – Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.
- Amazon EMR fügt Unterstützung hinzu, sodass Sie andere Anwendungen wie HBase Flink und Hive mit der Amazon S3 Express One Zone-Speicherklasse verwenden können.
- Diese Version bietet die Möglichkeit, wiederhergestellte Objekte zu lesen, sodass Sie Glacier-Objekte mit dem Protokoll von einem S3-Standort aus lesen können. S3A Diese Funktion funktioniert mit Spark, Flink und Hive.
- Node-Labels mit verwalteter Skalierung — Verwenden Sie Node-Labels für Ihre Instances, die auf dem Markttyp oder dem Knotentyp basieren, um die automatische Skalierung von Amazon zu verbessernEMR. Weitere Informationen finden Sie unter [Managed Scaling in Amazon verwenden EMR](#).

Bekannte Probleme

- Python 3.11 wird von EMR Studio nicht unterstützt.

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt ein Deadlock-Problem, das bei internen Cleanup-Vorgängen auftreten kann. Dieser Vorgang verwaltet den Lebenszyklus der einzelnen Schritte, während sie auf dem EMR Cluster abgeschlossen werden. Dieses Problem betrifft kritische EMR Amazon-Abläufe wie den schrittweisen Betrieb und die Skalierung.
- Diese Version behebt ein Problem, bei dem benutzerdefinierte Cluster mit benutzerdefinierten Clustern, die bereits über bestimmte Protokolldateien verfügenAMIs, dazu führen können, dass der Amazon EMR Log Management Daemon fehlschlägt.
- Amazon EMR 7.2.0 aktualisiert den EMR Amazon-Daemon, der für die Clusterverwaltung und die Überwachungsaktivitäten verantwortlich ist, von AWS SDK Version 1 auf Version 2.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amt Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.5.2 240708.1	6.1.96-102.177.amzn2023	8. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

Amazon EMR 6.15.0 (neueste Version der 6.x-Serie)

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.15.0. Änderungen beziehen sich auf Version 6.14.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.15.0 Änderungsprotokoll](#).

Neue Features

- Anwendungs-Upgrades – 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.
- [Schnellere Starts für EMR Cluster, die auf einem Cluster laufen EC2](#) — Es ist jetzt bis zu 35% schneller, einen EMR EC2 Amazon-On-Cluster zu starten. Mit dieser Verbesserung können die meisten Kunden ihre Cluster in 5 Minuten oder weniger starten.
- [CodeWhisperer für EMR Studio](#) — Sie können Amazon jetzt CodeWhisperer mit Amazon EMR Studio verwenden, um Empfehlungen in Echtzeit zu erhalten, während Sie Code einschreiben JupyterLab. CodeWhisperer kann Ihre Kommentare vervollständigen, einzelne Codezeilen fertigstellen, line-by-line Empfehlungen aussprechen und vollständig formatierte Funktionen generieren.
- [Schnellere Job-Neustartzeiten mit Flink](#) — Mit Amazon EMR 6.15.0 und höher sind mehrere neue Mechanismen für Apache Flink verfügbar, um die Job-Neustartzeit bei

Aufgabenwiederherstellungs- oder Skalierungsvorgängen zu verbessern. Dadurch wird die Geschwindigkeit der Wiederherstellung und des Neustarts von Ausführungsdiagrammen optimiert, um die Auftragsstabilität zu verbessern.

- Detaillierte [Zugriffskontrolle auf Tabellenebene für Open-Table-Formate](#) — Wenn Sie mit Amazon EMR 6.15.0 und höher Spark-Jobs auf Amazon auf EC2 Clustern ausführen, die auf Daten im AWS Glue-Datenkatalog zugreifen, können Sie damit Berechtigungen EMR auf Tabellen-, Zeilen-, Spalten- und Zellenebene auf Hudi-, Iceberg- oder Delta Lake-basierte Tabellen anwenden. AWS Lake Formation
- Hadoop-Upgrade — Amazon EMR 6.15.0 beinhaltet ein Upgrade von Apache Hadoop auf Version 3.3.6. Hadoop 3.3.6 war die neueste Version zum Zeitpunkt der Bereitstellung von Amazon EMR 6.15, die im Juni 2023 von Apache veröffentlicht wurde. Frühere Versionen von Amazon EMR (6.9.0 bis 6.14.x) verwendeten Hadoop 3.3.3.

Das Upgrade umfasst Hunderte von Verbesserungen und Fehlerbehebungen sowie Funktionen wie rekonfigurierbare Datenknotenparameter, die DFSAdmin Option zum Initiieren von Massenrekonfigurationsvorgängen an allen Live-Datenknoten und eine API Vektorfunktion, die es leseintensiven Lesern ermöglicht, mehrere Lesebereiche anzugeben. Hadoop 3.3.6 bietet außerdem Unterstützung HDFS APIs und Semantik für sein Write-Ahead-Protokoll (`WAL`), sodass es auf anderen Speichersystemimplementierungen ausgeführt werden kann. Weitere Informationen finden Sie in den Änderungsprotokollen für Versionen [3.3.4](#), [3.3.5](#) und [3.3.6](#) in der Apache-Hadoop-Dokumentation.

- Support AWS SDK für Java, Version 2 — Amazon EMR 6.15.0-Anwendungen können AWS SDK für die Java-Versionen [1.12.569](#) oder [2.20.160](#) verwendet werden, wenn die Anwendung v2 unterstützt. Die AWS SDK für Java 2.x ist eine grundlegende Neufassung der Codebasis von Version 1.x. Sie basiert auf Java 8+ und fügt mehrere häufig angeforderte Funktionen hinzu. Dazu gehören die Unterstützung für blockierungsfreie I/O und die Möglichkeit, zur Laufzeit eine andere HTTP Implementierung einzubinden. Weitere Informationen, einschließlich eines Migrationshandbuchs von SDK Java v1 zu v2, finden Sie im Handbuch [AWS SDK für Java, Version 2](#).

Änderungen, Verbesserungen und behobene Probleme

- Um Ihre EMR Hochverfügbarkeitscluster zu verbessern, ermöglicht diese Version Konnektivität zu EMR Amazon-Daemons auf lokalen Hosts, die Endpunkte verwenden. IPv6
- Diese Version ermöglicht TLS 1.2 für die Kommunikation mit ZooKeeper Provisioned auf allen primären Knoten Ihres Hochverfügbarkeitsclusters.

- Diese Version verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und den Clusterbetrieb unterbrechen.
- Diese Version macht die knoteninterne Kommunikation für Hochverfügbarkeitscluster widerstandsfähiger. EMR Durch diese Verbesserung wird die Wahrscheinlichkeit, dass Bootstrap-Aktionen oder Cluster-Startfehler auftreten, verringert.
- Tez in Amazon EMR 6.15.0 führt Konfigurationen ein, die Sie angeben können, um die Eingabe-Splits asynchron in einem Tez-gruppiereten Split zu öffnen. Dies führt zu einer schnelleren Leistung von Leseabfragen, wenn ein einzelner gruppierter Tez-Split eine große Anzahl von Eingabe-Splits enthält. Weitere Informationen finden Sie unter [Asynchrone Öffnung von Tez-Splits](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amaon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 206.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	13. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

Amazon EMR 5.36.2 (neueste Version der 5.x-Serie)

Neue EMR Amazon-Versionen werden in verschiedenen Regionen über einen Zeitraum von mehreren Tagen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.36.2. Die Änderungen beziehen sich auf 5.36.1. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Amazon die Kernknoten EMR nicht unter die für den Cluster eingestellte HDFS Replikationsfaktor-Einstellung herunterskaliert. Diese Verbesserung erfüllt die Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für die Ausführung Presto oder Trino einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung reduziert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit ausgeführt wird. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Behebt ein Problem, bei dem Cluster-Scale-Down-Operationen möglicherweise zum Stillstand kommen, während Amazon EMR einen Core-Node ordnungsgemäß außer Betrieb nimmt und dieser vor der vollständigen Außerbetriebnahme fehlerhaft wird.
- Verbessert die Stabilität eines Knotens in einem Hochverfügbarkeitscluster mit mehreren Primärknoten, wenn Amazon einen einzelnen Knoten EMR neu startet.

- Optimiert die Protokollverwaltung, wenn Amazon auf Amazon EMR EC2 läuft. Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und den Clusterbetrieb unterbrechen.
- Behebt einen seltenen Fehler, der dazu führen kann, dass ein Hochverfügbarkeitscluster mit mehreren Primärknoten ausfällt, weil er nicht mit dem Yarn kommunizieren kann.
ResourceManager
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 503.0	4,14.343	xxxxxx, 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

Amazon EMR und AWS Signature Version 4

EMR Amazon-Versionen verwenden AWS Signature Version 4 (SigV4), um Anfragen an Amazon S3 zu authentifizieren. Buckets, die nach dem 24. Juni 2020 in Amazon S3 erstellt wurden, unterstützen keine Anfragen, die mit Signature Version 2 (SIGv2) signiert wurden. Buckets, die am oder vor dem 24. Juni 2020 erstellt wurden, unterstützen weiterhin SIGv2. Wir empfehlen Ihnen, zu einer EMR Amazon-Version zu migrieren, die SigV4 unterstützt, damit Sie auf neue S3-Buckets zugreifen und Unterbrechungen Ihrer Workloads vermeiden können.

Wenn Sie Anwendungen verwenden, die in Amazon enthalten sind, EMR wie Apache Spark, Apache Hive und Presto, müssen Sie Ihren Anwendungscode nicht ändern, um ihn zu verwenden. SigV4 Wenn Sie benutzerdefinierte Anwendungen verwenden, die nicht in Amazon enthalten sind, müssen Sie möglicherweise Ihren Code aktualisieren, um ihn verwenden zu können SigV4. Weitere Informationen finden Sie unter [Umstieg von Signature Version 2 auf Signature Version 4](#) im Amazon-S3-Benutzerhandbuch.

Die folgenden EMR Amazon-Versionen unterstützen 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. Alle 6.x- und 7.x-Versionen unterstützen SigV4.

Ansatz zur Minderung CVE -2021-44228

Note

Für EMR Amazon-Version 6.9.0 und höher verwenden alle von Amazon installierten Komponenten, EMR die Log4j-Bibliotheken verwenden, Log4j Version 2.17.1 oder höher.

Amazon EMR läuft weiter EC2

Das in [CVE-2021-44228](#) behandelte Problem ist für Apache Log4j-Kernversionen zwischen 2.0.0 und 2.14.1 relevant, wenn Eingaben aus nicht vertrauenswürdigen Quellen verarbeitet werden. EMRAmazon-Cluster, die mit Amazon EMR 5.x-Versionen bis 5.34.0 und EMR 6.x-Versionen bis Amazon EMR 6.5.0 gestartet wurden, enthalten Open-Source-Frameworks wie Apache Hive, Flink,, Presto und Trino, die diese Versionen von Apache Log4j verwenden. HUDI Viele Kunden verwenden jedoch die auf ihren EMR Amazon-Clustern installierten Open-Source-Frameworks, um Eingaben aus nicht vertrauenswürdigen Quellen zu verarbeiten und zu protokollieren.

Wir empfehlen Ihnen, die „Amazon EMR Bootstrap Action Solution for Log4j CVE -2021-44228“ wie im folgenden Abschnitt beschrieben anzuwenden. Diese Lösung adressiert auch -2021-45046. CVE

Note

Die Bootstrap-Aktionsskripte für Amazon EMR wurden am 7. September 2022 aktualisiert und enthalten nun schrittweise Bugfixes und Verbesserungen für Oozie. Wenn Sie Oozie verwenden, sollten Sie die aktualisierte Amazon EMR Bootstrap Action-Lösung anwenden, die im folgenden Abschnitt beschrieben wird.

Amazon EMR auf EKS

Wenn Sie [Amazon EMR on EKS mit der](#) Standardkonfiguration verwenden, sind Sie von dem in CVE -2021-44228 beschriebenen Problem nicht betroffen, und Sie müssen die im Abschnitt beschriebene Lösung nicht anwenden. [Amazon EMR Bootstrap-Aktionslösung für Log4j CVE -2021-44228 & -2021-45046 CVE](#) Für Amazon EMR on EKS verwendet die EMR Amazon-Runtime für Spark Apache Log4j Version 1.2.17. Wenn Sie Amazon EMR on verwenden, sollten EKS Sie die Standardeinstellung für die `log4j . appender` Komponente nicht auf `ändernLog`.

Amazon EMR Bootstrap-Aktionslösung für Log4j CVE -2021-44228 & -2021-45046 CVE

Diese Lösung bietet eine EMR Amazon-Bootstrap-Aktion, die auf Ihre EMR Amazon-Cluster angewendet werden muss. Für jede EMR Amazon-Version finden Sie unten einen Link zu einem Bootstrap-Aktionsskript. Sie sollten die folgenden Schritte ausführen, um diese Bootstrap-Aktion anzuwenden:

1. Kopieren Sie das Skript, das Ihrer EMR Amazon-Version entspricht, in einen lokalen S3-Bucket in Ihrem AWS-Konto. Bitte stellen Sie sicher, dass Sie ein Bootstrap-Skript verwenden, das für Ihre EMR Amazon-Version spezifisch ist.
2. [Richten Sie eine Bootstrap-Aktion für Ihre EMR Cluster ein, um das in Ihren S3-Bucket kopierte Skript gemäß den in EMR der Dokumentation beschriebenen Anweisungen auszuführen.](#) Wenn Sie andere Bootstrap-Aktionen für Ihre EMR Cluster konfiguriert haben, stellen Sie bitte sicher, dass dieses Skript als erstes Bootstrap-Aktionsskript eingerichtet ist, das ausgeführt wird.
3. Beenden Sie vorhandene EMR Cluster und starten Sie neue Cluster mit dem Bootstrap-Aktionsskript. AWS empfiehlt, dass Sie die Bootstrap-Skripte in Ihrer Testumgebung testen und Ihre Anwendungen validieren, bevor Sie sie in Ihrer Produktionsumgebung anwenden. Wenn Sie nicht die neueste Version für eine EMR Nebenversion (z. B. 6.3.0) verwenden, müssen Sie die neueste Version (z. B. 6.3.1) verwenden und dann die oben beschriebene Lösung anwenden.

CVE-2021-44228 & CVE -2021-45046 - Bootstrap-Skripte für Amazon-Releases EMR

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
6.5.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.5.0-v2.sh</code>	24. März 2022
6.4.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.4.0-v2.sh</code>	24. März 2022

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
6.3.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.3.1-v2.sh</code>	24. März 2022
6.2.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.2.1-v2.sh</code>	24. März 2022
6.1.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.1.1-v2.sh</code>	14. Dezember 2021
6.0.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.0.1-v2.sh</code>	14. Dezember 2021
5,34,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.34.0-v2.sh</code>	12. Dezember 2021
5,33,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.33.1-v2.sh</code>	12. Dezember 2021

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5.32,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.32.1-v2.sh</code>	13. Dezember 2021
5.31,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.31.1-v2.sh</code>	13. Dezember 2021
5,30,2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.30.2-v2.sh</code>	14. Dezember 2021
5.29,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.29.0-v2.sh</code>	14. Dezember 2021
5,28,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.28.1-v2.sh</code>	15. Dezember 2021
5,27,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.27.1-v2.sh</code>	15. Dezember 2021

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5,26,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.26.0-v2.sh</code>	15. Dezember 2021
5,25,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.25.0-v2.sh</code>	15. Dezember 2021
5,24,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.24.1-v2.sh</code>	15. Dezember 2021
5.23,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.23.1-v2.sh</code>	15. Dezember 2021
5,22,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.22.0-v2.sh</code>	15. Dezember 2021
5,21,2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.21.2-v2.sh</code>	15. Dezember 2021

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5.20,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.20.1-v2.sh</code>	15. Dezember 2021
5.19,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.19.1-v2.sh</code>	15. Dezember 2021
5.18.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.18.1-v2.sh</code>	15. Dezember 2021
5.17.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.17.2-v2.sh</code>	15. Dezember 2021
5.16.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.16.1-v2.sh</code>	15. Dezember 2021
5.15,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.15.1-v2.sh</code>	15. Dezember 2021

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5.14.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.14.2-v2.sh</code>	15. Dezember 2021
5.13.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.13.1-v2.sh</code>	15. Dezember 2021
5.12.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.12.3-v2.sh</code>	15. Dezember 2021
5.11.4	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.11.4-v2.sh</code>	15. Dezember 2021
5.10.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.10.1-v2.sh</code>	15. Dezember 2021
5.9.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.9.1-v2.sh</code>	15. Dezember 2021

EMRAmazon-Veröffentlichungsnummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5.8.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.8.3-v2.sh</code>	15. Dezember 2021
5.7.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.7.1-v2.sh</code>	15. Dezember 2021

EMRVersion veröffentlichen	Letzte Version vom Dezember 2021
6.3.0	6.3.1
6.2.0	6.2.1
6.1.0	6.1.1
6.0.0	6.0.1
5.33,0	5.33,1
5,32,0	5,32,1
5,31,0	5,31,1
5.30.0 oder 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

EMRVersion veröffentlichen	Letzte Version vom Dezember 2021
5.21.0 oder 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 oder 5.17.1	5.17.2
5.16.0	5.16.1
5.15.0	5.15,1
5.14.0 oder 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

Häufig gestellte Fragen

- Sind EMR Versionen, die älter als EMR 5 sind, von -2021-44228 betroffen? CVE

Nein. EMRVersionen vor EMR Version 5 verwenden Log4j-Versionen, die älter als 2.0 sind.

- Adressiert diese Lösung -2021-45046CVE?

Ja, diese Lösung adressiert auch -2021-45046. CVE

- Unterstützt die Lösung benutzerdefinierte Anwendungen, die ich auf meinen Clustern installiere?
EMR

Das Bootstrap-Skript aktualisiert nur JAR Dateien, die von installiert wurden EMR. Wenn Sie benutzerdefinierte Anwendungen und JAR Dateien auf Ihren EMR Clustern über Bootstrap-Aktionen, als an Ihre Cluster übermittelte Schritte, mithilfe von benutzerdefiniertem Amazon Linux AMI oder über einen anderen Mechanismus installieren und ausführen, ermitteln Sie gemeinsam mit Ihrem Anwendungsanbieter, ob Ihre benutzerdefinierten Anwendungen von CVE -2021-44228 betroffen sind, und suchen Sie nach einer geeigneten Lösung.

- Wie sollte ich mit [benutzerdefinierten Docker-Images](#) mit EMR on EKS umgehen?

Wenn Sie benutzerdefinierte Anwendungen EKS mithilfe benutzerdefinierter [Docker-Images](#) zu EMR Amazon hinzufügen oder Jobs mit EKSwith benutzerdefinierten Anwendungsdateien EMR an Amazon senden, ermitteln Sie gemeinsam mit dem Anbieter der Anwendung, ob Ihre benutzerdefinierten Anwendungen von CVE -2021-44228 betroffen sind, und ermitteln Sie eine geeignete Lösung.

- Wie funktioniert das Bootstrap-Skript, um das in -2021-44228 und -2021-45046 beschriebene Problem zu beheben? CVE CVE

Das Bootstrap-Skript aktualisiert die Startanweisungen, indem es einen neuen Befehlssatz EMR hinzufügt. Diese neuen Anweisungen löschen die JndiLookup Klassendateien, die über Log4j von allen Open-Source-Frameworks verwendet werden, die von installiert wurden. EMR Dies folgt der [von Apache veröffentlichten Empfehlung](#) zur Behebung der Log4j-Probleme.

- Gibt es ein Update EMR, das Log4j-Versionen 2.17.1 oder höher verwendet?

EMR5 Releases bis Version 5.34 und EMR 6 Releases bis Version 6.5 verwenden ältere Versionen von Open-Source-Frameworks, die mit den neuesten Versionen von Log4j nicht kompatibel sind. Wenn Sie diese Versionen weiterhin verwenden, empfehlen wir Ihnen, die Bootstrap-Aktion anzuwenden, um die in der beschriebenen Probleme zu beheben. CVEs Nach Version EMR 5 5.34 und EMR 6 Version 6.5 werden Anwendungen, die Log4j 1.x und Log4j 2.x verwenden, auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert, sodass die oben genannten Bootstrap-Aktionen nicht erforderlich sind, um die CVE Probleme zu beheben.

- Sind EMR Versionen von -2021-45105 betroffen? CVE

Die von Amazon EMR mit EMR den Standardkonfigurationen installierten Anwendungen sind von CVE -2021-45105 nicht betroffen. Unter den von Amazon EMR installierten Anwendungen verwendet nur Apache Hive Apache Log4j mit [Kontext-Lookups](#) und verwendet kein vom Standard abweichendes Muster-Layout, sodass unangemessene Eingabedaten verarbeitet werden können.

- Ist Amazon von einer der folgenden CVE Offenlegungen EMR betroffen?

Die folgende Tabelle enthält eine Liste von CVEs, die sich auf Log4j beziehen, und gibt an, ob sie sich jeweils CVE auf Amazon auswirken. EMR Die Informationen in dieser Tabelle gelten nur, wenn Anwendungen von Amazon EMR mit den Standardkonfigurationen installiert werden.

CVE	Auswirkungen EMR	Hinweise
CVE-2022-23302	Nein	Amazon EMR richtet Log4j nicht ein JMSSink
CVE-2022-23305	Nein	Amazon EMR richtet Log4j nicht ein JDBCAppender
CVE-2022-23307	Nein	Amazon EMR richtet Log4j Chainsaw nicht ein
CVE-2020-9493	Nein	Amazon EMR richtet Log4j Chainsaw nicht ein
CVE-2021-44832	Nein	Amazon EMR richtet Log4j nicht JDBCAppender mit einer Verbindungszeichenfolge ein JNDI
CVE-2021-4104	Nein	Amazon verwendet Log4j EMR nicht JMSAppender
CVE-2020-9488	Nein	Die von Amazon installierten Anwendungen verwenden Log4j EMR nicht SMTPAppender
CVE-2019-17571	Nein	Amazon EMR blockiert den öffentlichen Zugriff auf Cluster und startet nicht SocketServer

CVE	Auswirkungen EMR	Hinweise
CVE-2019-17531	Nein	Wir empfehlen Ihnen, auf die neueste EMR Amazon-Release-Version zu aktualisieren. Amazon EMR 5.33.0 und höher verwenden jackson-databind 2.6.7.4 oder höher, und EMR 6.1.0 und höher verwenden jackson-databind 2.10.0 oder höher. Diese Versionen von jackson-databind sind von dem nicht betroffen. CVE

EMRAmazon-Archiv mit Versionshinweisen

Versionshinweise für alle EMR Amazon-Versionen sind unten verfügbar. Umfassende Versionsinformationen für jede Version finden Sie unter [Release-Versionen von Amazon EMR 6.x](#), [Amazon EMR 5.x-Release-Versionen](#) und [Amazon EMR 4.x-Release-Versionen](#).

Um Updates zu erhalten, wenn eine neue EMR Amazon-Version verfügbar ist, abonnieren Sie den [RSSFeed für EMR Amazon-Versionshinweise](#).

Version 6.14.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.14.0. Änderungen beziehen sich auf Version 6.13.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.14.0 Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- [Amazon EMR Managed Scaling](#) ist jetzt in der Region ap-southeast-3 Asien-Pazifik (Jakarta) für Cluster verfügbar, die Sie mit Amazon EMR 6.14.0 und höher erstellen.

Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.14.0 optimiert die Protokollverwaltung, wenn Amazon auf Amazon EMR läuft. EC2 Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Die Version 6.14.0 verbessert den Skalierungs-Workflow, um verschiedene Core-Instances zu berücksichtigen, deren Größe für ihre EBS Amazon-Volumes erheblich variiert. Diese Verbesserung gilt nur für Core-Knoten; Herunterskalierungs-Operationen für Aufgabenknoten sind davon nicht betroffen.
- Die Version 6.14.0 verbessert die Art und Weise, wie Amazon mit Open-Source-Anwendungen wie EMR interagiert. Apache Hadoop YARN ResourceManager and HDFS NameNode Diese Verbesserung reduziert das Risiko von Betriebsverzögerungen bei der Cluster-Skalierung und verringert Startup-Fehler, die aufgrund von Verbindungsproblemen mit den Open-Source-Anwendungen auftreten.
- Die Version 6.14.0 optimiert die Anwendungsinstallation beim Clusterstart. Dies verbessert die Cluster-Startzeiten für bestimmte Kombinationen von EMR Amazon-Anwendungen.
- Die Version 6.14.0 behebt ein Problem, bei dem Cluster-Scale-Down-Operationen zum Stillstand kommen können, wenn ein Cluster, der in einer VPC mit einer benutzerdefinierten Domain ausgeführt wird, auf einen Core- oder Task-Node-Neustart stößt.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	17. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	11. September 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

Version 6.13.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.13.0.

Änderungen beziehen sich auf Version 6.12.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.13.0 Änderungsprotokoll](#).

Neue Features

- 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.

Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.13.0 verbessert den Amazon EMR Log Management Daemon, um sicherzustellen, dass alle Protokolle in regelmäßigen Abständen auf Amazon S3 hochgeladen werden, wenn ein Befehl zur Clusterbeendigung ausgegeben wird. Dies ermöglicht schnellere Clusterbeendigungen.
- Die Version 6.13.0 erweitert die EMR Amazon-Protokollverwaltungsfunktionen, um einen konsistenten und zeitnahen Upload aller Protokolldateien auf Amazon S3 sicherzustellen. Davon profitieren vor allem Cluster mit langer LaufzeitEMR.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 906.0	4,14.322	04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

Version 6.12.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.12.0.

Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.12.0 Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.12.0 unterstützt 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.
- Amazon EMR veröffentlicht 6.12.0 und höher und unterstützt die LDAP Integration mit Apache Livy, Apache Hive bis HiveServer 2 (HS2), Trino, Presto und Hue. Sie können Apache Spark und Apache Hadoop auch auf einem EMR Cluster installieren, der 6.12.0 oder höher verwendet, und sie für die Verwendung konfigurieren. LDAP Weitere Informationen finden Sie unter [Verwenden von Active Directory oder LDAP Servern für die Authentifizierung bei Amazon EMR](#).

Änderungen, Verbesserungen und behobene Probleme

- EMRAmazon-Versionen 6.12.0 und höher bieten Java 11-Laufzeitunterstützung für Flink. Weitere Informationen finden Sie unter [Flink für die Ausführung mit Java 11 konfigurieren](#).
- Die Version 6.12.0 fügt dem Cluster-Skalierungs-Workflow für EMR Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Die Version 6.12.0 behebt ein Problem, bei dem Cluster-Herunterskalierungs-Operationen zum Stillstand kommen können, wenn ein Core-Knoten, der einer ordnungsgemäßen Außerbetriebnahme unterzogen wird, aus irgendeinem Grund nicht mehr funktionsfähig ist, bevor er vollständig außer Betrieb genommen wird.
- Die Version 6.12.0 verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter den für den Cluster eingestellten Replikationsfaktor herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Die Version 6.12.0 verbessert die Leistung und Effizienz des Health Monitoring Service für Amazon, EMR indem die Geschwindigkeit erhöht wird, mit der Statusänderungen für Instances

protokolliert werden. Diese Verbesserung verringert das Risiko einer Leistungseinbuße bei Cluster-Knoten, auf denen mehrere benutzerdefinierte Client-Tools oder Drittanbieteranwendungen ausgeführt werden.

- Die Version 6.12.0 verbessert die Leistung des On-Cluster-Log-Management-Daemons für Amazon EMR. Daher ist die Wahrscheinlichkeit eines Leistungseinbruchs bei EMR Clustern, die Schritte mit hoher Parallelität ausführen, geringer.
- Mit der EMR Amazon-Version 6.12.0 wurde der Log-Management-Daemon aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.
- Die Version 6.12.0 beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Die Version 6.12.0 ermöglicht die Protokollrotation für Timeline Server-Protokolle. YARN Dadurch werden Szenarien mit übermäßiger Festplattenauslastung minimiert, insbesondere bei Clustern mit langer Laufzeit.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amt-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 906.0	4,14.322	04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

Version 6.11.1

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.11.1.

Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.11.1 Änderungsprotokoll](#).

Änderungen, Verbesserungen und behobene Probleme

- Aufgrund von Sperrkonflikten kann ein Knoten in einen Deadlock geraten, wenn er gleichzeitig mit dem Versuch, ihn außer Betrieb zu nehmen, hinzugefügt oder entfernt wird. Infolgedessen reagiert der Hadoop Resource Manager (YARN) nicht mehr und wirkt sich auf alle eingehenden und aktuell laufenden Container aus.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version enthält Sicherheitskorrekturen für Hue und HBase
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon ausführen, EMR möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith` und `endsWith` erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn keine benutzerdefinierten Funktionen (UDF) vorhanden sind.
- Diese Version behebt ein Problem, bei dem Containerprotokolle vom Node Log Aggregation Service gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme an S3 weiterleiten kann.
- Diese Version behebt ein Problem mit FairShare Scheduler-Metriken, wenn Node Label für Hadoop aktiviert ist.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version erweitert den Cluster-Skalierungs-Workflow für EMR Cluster, auf denen Presto oder Trino ausgeführt wird, um einen neuen Wiederholungsmechanismus. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen

Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.

- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Ihr Cluster nicht versucht, die Kernknoten unter die für den Cluster eingestellte Replikationsfaktor-Einstellung herunterzuskalieren. HDFS Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon die Dateien EMR ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle auf Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der EMR Amazon-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Paz

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			ifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906.0		04. Oktober 2023	China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

Version 6.11.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.11.0. Änderungen beziehen sich auf Version 6.10.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.11.0 unterstützt Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg 1.2.0-amzn-0, Trino 410-amzn-0 und PrestoDB 0.279-amzn-0.

Änderungen, Verbesserungen und behobene Probleme

- Mit Amazon EMR 6.11.0 wurde der DynamoDB-Connector auf Version 5.0.0 aktualisiert. Version 5.0.0 verwendet AWS SDK for Java 2.x In früheren Versionen wurde AWS SDK for Java 1.x verwendet. Aufgrund dieses Upgrades empfehlen wir Ihnen dringend, Ihren Code zu testen, bevor Sie den DynamoDB-Connector mit Amazon EMR 6.11 verwenden.
- Wenn der DynamoDB-Connector für Amazon EMR 6.11.0 den DynamoDB-Service aufruft, verwendet er den Region-Wert, den Sie für die Eigenschaft angeben. `dynamodb.endpoint` Wir empfehlen, dass Sie auch `dynamodb.region` konfigurieren, wenn Sie `dynamodb.endpoint` verwenden, und dass beide Eigenschaften auf dieselbe AWS-Region abzielen. Wenn Sie verwenden `dynamodb.endpoint` und nicht konfigurieren `dynamodb.region`, gibt der DynamoDB-Connector für Amazon EMR 6.11.0 eine ungültige Region-Ausnahme zurück und versucht, Ihre AWS-Region Informationen aus dem EC2 Amazon-Instance-Metadatenservice (IMDS) abzugleichen. Wenn der Connector die Region nicht abrufen kann, verwendet er standardmäßig USA Ost (Nord-Virginia) (`us-east-1`). Der folgende Fehler ist ein Beispiel für die ungültige Region-Ausnahme, die möglicherweise angezeigt wird, wenn Sie die `dynamodb.region` Eigenschaft nicht richtig konfigurieren: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region`. Weitere Informationen zu den Klassen, die vom AWS SDK for Java Upgrade auf 2.x betroffen sind, finden Sie im Commit [Upgrade AWS SDK for Java from 1.x auf 2.x \(#175\)](#) im GitHub Repository für den EMR Amazon-DynamoDB-Connector.
- Diese Version behebt ein Problem, bei dem Spaltendaten zu NULL werden, wenn Sie Delta Lake zum Speichern von Delta-Tabellendaten in Amazon S3 nach dem Spaltenumbenennungsvorgang verwenden. Weitere Informationen zu diesem experimentellen Feature in Delta Lake finden Sie unter [Vorgang zum Umbenennen von Spalten](#) im Delta-Lake-Benutzerhandbuch.
- Die Version 6.11.0 behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen


eines Edge-Knotens für die Kommunikation mit Ihrem EMR Cluster finden Sie im Repo unter [Edge Node Creator](#). [aws-samples](#) GitHub

- Die Version 6.11.0 verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um EBS Amazon-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Die Version 6.11.0 behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon auf Amazon EMR veröffentlicht. [CloudWatch](#)
- Die Version 6.11.0 behebt ein Problem mit EMR Clustern, bei dem eine Aktualisierung der YARN Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer übermäßigen Festplattenauslastung unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container ausgeführt wurden, so lange außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` ist der Wert von `true` auf `false` in `yarn-site.xml` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN -9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Jobs aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit EMR Amazon-Versionen 6.11.0 und höher verfügbar.

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

 Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6 . 8 . **1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API Loder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504,1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

Version 6.10.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.10.0.

Änderungen beziehen sich auf Version 6.9.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Neue Features

- Amazon EMR 6.10.0 unterstützt 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 und PrestoDB 0.278.1.
- Amazon EMR 6.10.0 enthält einen nativen Trino-Hudi-Konnektor, der Lesezugriff auf Daten in Hudi-Tabellen ermöglicht. Sie können den Konnektor mit `trino-cli --catalog hudi` aktivieren und den Konnektor für Ihre Anforderungen mit `trino-connector-hudi` konfigurieren. EMRDurch die native Integration mit Amazon müssen Sie Hudi-Tabellen nicht mehr abfragen. `trino-connector-hive` Eine Liste der unterstützten Konfigurationen mit dem neuen Konnektor finden Sie auf der [Hudi-Konnektor-Seite](#) der Trino-Dokumentation.
- EMRAmazon-Versionen 6.10.0 und höher unterstützen die Apache Zeppelin-Integration mit Apache Flink. Weitere Informationen finden Sie unter [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#).

Bekannte Probleme

- Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Um dieses Problem in Amazon EMR 6.10.0 zu umgehen, können Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` setzen. In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration `false` standardmäßig auf eingestellt, um dieses Problem zu beheben.

Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR 6.10.0 entfernt die Abhängigkeit von der [Amazon Redshift Redshift-Integration minimal-json.jar für Apache Spark](#) und fügt automatisch die erforderlichen Spark-Redshift-bezogenen JAR-Dateien zum Executor-Klassenpfad für Spark hinzu: `und. spark-redshift.jar` `spark-avro.jar` `RedshiftJDBC.jar`
- Die Version 6.10.0 verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem Cluster. EMR Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Die Version 6.10.0 startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Amazon EMR 6.10.0 unterstützt regionale Endpunkte für die EMRFS Benutzerzuweisung.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Die Version 6.10.0 behebt ein Problem, das dazu führte, dass Spark-Jobs zum Stillstand kamen, wenn sich alle verbleibenden Spark-Executoren auf einem außer Betrieb genommenen Host mit dem Resource Manager befanden. YARN
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf `-1` zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version

verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API Loder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504,1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Arta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404,1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)
2.0.2023 404,0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE)

Version 6.9.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.9.0. Die Änderungen beziehen sich auf EMR Amazon-Version 6.8.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Neue Features

- EMRAmazon-Version 6.9.0 unterstützt Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 und Tez 0.10.2.
- EMRAmazon-Version 6.9.0 enthält eine neue Open-Source-Anwendung, [Delta Lake](#) 2.1.0.
- Die Amazon Redshift Redshift-Integration für Apache Spark ist in EMR Amazon-Versionen 6.9.0 und höher enthalten. Die native Integration war bisher ein Open-Source-Tool und ist ein Spark-Konnektor, mit dem Sie Apache-Spark-Anwendungen erstellen können, die Daten in Amazon Redshift und Amazon Redshift Serverless lesen und in diese schreiben. Weitere Informationen finden Sie unter [Verwenden der Amazon Redshift Redshift-Integration für Apache Spark mit Amazon EMR](#).
- Die EMR Amazon-Version 6.9.0 bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Scale-Down. Bisher konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die neue Funktion stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).
- Um Abfragen mit langer Laufzeit zu unterstützen, verfügt Trino jetzt über einen fehlertoleranten Ausführungsmechanismus. Die fehlertolerante Ausführung minimiert Abfragefehler, indem fehlgeschlagene Abfragen oder deren Komponentenaufgaben wiederholt werden. Weitere Informationen finden Sie unter [Fehlertolerante Ausführung in Trino](#).
- Sie können Apache Flink on Amazon EMR für die Vereinheitlichung BATCH und STREAM Verarbeitung von Apache Hive-Tabellen oder Metadaten beliebiger Flink-Tabellenquellen wie Iceberg, Kinesis oder Kafka verwenden. Sie können den AWS Glue-Datenkatalog als Metastore für Flink angeben, indem Sie Amazon AWS Management Console, AWS CLI, oder verwenden. EMR API Weitere Informationen finden Sie unter [Konfiguration von Flink in Amazon EMR](#).
- Sie können jetzt AWS Identity and Access Management (IAM) Runtime-Rollen und AWS Lake Formation basierte Zugriffskontrolle für Apache Spark-, Apache Hive- und Presto-Abfragen auf Amazon EMR auf EC2 Clustern mit Amazon SageMaker Studio angeben. Weitere Informationen finden [Sie unter EMR Schritte zur Konfiguration von Runtime-Rollen für Amazon](#).

Bekannte Probleme

- Für die EMR Amazon-Version 6.9.0 funktioniert Trino nicht auf Clustern, die für Apache Ranger aktiviert sind. Wenn Sie Trino mit Ranger verwenden müssen, wenden Sie sich an [AWS Support](#).

- Wenn Sie die Amazon-Redshift-Integration für Apache Spark verwenden und eine Zeit, `timetz`, `timestamp` oder `timestampz` mit Mikrosekundengenauigkeit im Parquet-Format haben, rundet der Konnektor die Zeitwerte auf den nächstliegenden Millisekundenwert. Um das Problem zu umgehen, verwenden Sie den `unload_s3_format`-Formatparameter-Text-Unload.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Verbindungen zu EMR Amazon-Clustern von Amazon SageMaker Studio aus können zeitweise fehlschlagen und der Antwortcode 403 Forbidden wird angezeigt. Dieser Fehler tritt auf, wenn die Einrichtung der IAM Rolle im Cluster länger als 60 Sekunden dauert. Um das Problem zu umgehen, können Sie einen EMR Amazon-Patch installieren, um Wiederholungsversuche zu ermöglichen und das Timeout auf mindestens 300 Sekunden zu erhöhen. Verwenden Sie die folgenden Schritte, um die Bootstrap-Aktion anzuwenden, wenn Sie Ihren Cluster starten.

1. Laden Sie das Bootstrap-Skript und die RPM Dateien vom folgenden Amazon S3 URIs herunter.

```
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. Laden Sie die Dateien aus dem vorherigen Schritt in einen Amazon-S3-Bucket hoch, dessen Eigentümer Sie sind. Der Bucket muss sich in dem Bereich befinden AWS-Region , in dem Sie den Cluster starten möchten.
3. Schließen Sie beim Starten Ihres EMR Clusters die folgende Bootstrap-Aktion ein. Ersetzen *bootstrap_URI* and *RPM_URI* mit dem entsprechenden URIs von Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei RecordServer Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. SecretAgent Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink bietet native S3 FileSystem - und FileSystem Hadoop-Konnektoren, mit denen Anwendungen Daten erstellen FileSink und in Amazon S3 schreiben können. Dies FileSink schlägt mit einer der folgenden zwei Ausnahmen fehl.

```
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:?]

```

Als Workaround können Sie einen EMR Amazon-Patch installieren, der das oben genannte Problem in Flink behebt. Führen Sie die folgenden Schritte aus, um die Bootstrap-Aktion beim Start Ihres Clusters anzuwenden.

1. Laden Sie das flink-rpm in Ihren Amazon-S3-Bucket herunter. Ihr RPM Weg ist. `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`
2. Laden Sie das Bootstrap-Skript und die RPM Dateien wie folgt URI von Amazon S3 herunter. `regionName` Ersetzen Sie es durch den AWS-Region Ort, an dem Sie den Cluster starten möchten.

```

s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh

```

3. Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierte verwaltete Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es eine Lösung für dieses Problem, indem Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` in den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration `false` standardmäßig auf eingestellt, um dieses Problem zu beheben.

Änderungen, Verbesserungen und behobene Probleme

- Für EMR Amazon-Version 6.9.0 und höher verwenden alle von Amazon installierten Komponenten, EMR die Log4j-Bibliotheken verwenden, Log4j Version 2.17.1 oder höher.
- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück,

obwohl der Eingabe-Split auf nicht leere Daten verweist. EMRAmazon-Version 6.9.0 behebt dieses Problem.

- Amazon EMR 6.9.0 bietet eingeschränkte Unterstützung für die auf Lake Formation basierende Zugriffskontrolle mit Apache Hudi beim Lesen von Daten mit Spark. SQL Die Unterstützung bezieht sich auf SELECT Abfragen, die Spark verwenden, SQL und ist auf die Zugriffskontrolle auf Spaltenebene beschränkt. Weitere Informationen finden Sie unter [Hudi und Lake Formation](#).
- Wenn Sie Amazon EMR 6.9.0 verwenden, um einen Hadoop-Cluster mit aktivierten [Node Labels](#) zu erstellen, geben die [YARNMetriken](#) aggregierte Informationen für alle Partitionen statt der Standardpartition API zurück. [Weitere Informationen finden Sie unter -11414. YARN](#)
- Mit der EMR Amazon-Version 6.9.0 haben wir Trino auf Version 398 aktualisiert, die Java 17 verwendet. Die vorherige unterstützte Version von Trino für Amazon EMR 6.8.0 war Trino 388, das auf Java 11 lief. Weitere Informationen zu dieser Änderung finden Sie unter [Trino-Updates für Java 17 im Trino-Blog](#).
- Diese Version behebt ein Problem mit der Nichtübereinstimmung der Timing-Sequenz zwischen Apache BigTop und EMR Amazon bei der EC2 Cluster-Startsequenz. Diese Nichtübereinstimmung der Timing-Sequenz tritt auf, wenn ein System versucht, zwei oder mehr Operationen gleichzeitig auszuführen, anstatt sie in der richtigen Reihenfolge auszuführen. Infolgedessen kam es bei bestimmten Clusterkonfigurationen zu Timeouts beim Instance-Start und zu langsameren Cluster-Startup-Zeiten.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Archiv-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI Aktion [ListReleaseLabels](#) API Loder. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia), Asien Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik (Melbourne), Israel (Tel Aviv), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	12. Januar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

Version 6.8.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.8.0. Änderungen beziehen sich auf Version 6.7.0.

Neue Features

- Die Amazon EMR Steps-Funktion unterstützt jetzt Apache Livy-Endpunkte und JDBC ODBC /-Clients. Weitere Informationen finden [Sie unter EMR Schritte zur Konfiguration von Runtime-Rollen für Amazon](#).
- EMRAmazon-Version 6.8.0 wird mit Apache HBase Version 2.4.12 geliefert. Mit dieser HBase Version können Sie Ihre Tabellen sowohl archivieren als auch löschen. HBase Der Amazon-S3-Archivierungsprozess benennt alle Tabellendateien in das Archivverzeichnis um. Dies kann ein kostspieliger und langwieriger Prozess sein. Jetzt können Sie den Archivierungsvorgang überspringen und große Tabellen schnell löschen und löschen. Weitere Informationen finden Sie unter [Verwenden der HBase Shell](#).

Bekannte Probleme

- Hadoop 3.3.3 führte eine Änderung in YARN ([YARN-9608](#)) ein, die dafür sorgt, dass Knoten, auf denen Container liefen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierte verwaltete Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es eine Lösung für dieses Problem, indem Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` In den EMR Amazon-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration `false` standardmäßig auf eingestellt, um dieses Problem zu beheben.

Änderungen, Verbesserungen und behobene Probleme

- Als Amazon EMR Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell las, produzierte Amazon EMR eine. `NoSuchMethodError` EMRAmazon-Version 6.8.0 behebt dieses Problem.
- Die EMR Amazon-Version 6.8.0 enthält [Apache Hudi](#) 0.11.1. Amazon EMR 6.8.0-Cluster sind jedoch auch mit der Open-Source-Version von Hudi 0.12.0 kompatibel. `hudi-spark3.3-bundle_2.12`

- EMRAmazon-Version 6.8.0 wird mit Apache Spark 3.3.0 geliefert. Diese Spark-Version verwendet Apache Log4j 2 und die `log4j2.properties`-Datei zur Konfiguration von Log4j in Spark-Prozessen. Wenn Sie Spark im Cluster verwenden oder EMR Cluster mit benutzerdefinierten Konfigurationsparametern erstellen und ein Upgrade auf EMR Amazon-Version 6.8.0 durchführen möchten, müssen Sie auf die neue `spark-log4j2` Konfigurationsklassifizierung und das neue Schlüsselformat für Apache Log4j 2 migrieren. Weitere Informationen finden Sie unter [Migration von Apache Log4j 1.x zu Log4j 2.x](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Architektur. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

Note

Diese Version erhält keine automatischen AMI Updates mehr, da ihr eine weitere Patch-Version folgte. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie das EMRAmazon-Release-Drop-down-Menü, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [list-release-labels](#) CLI-Aktion [ListReleaseLabels](#) API-Operation. Um über neue Versionen informiert zu werden, abonnieren Sie den RSS Feed unter [Was ist neu?](#) Seite.

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE), Europa (Spanien), Europa (Zürich), Asien-Pazifik

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Melbourne), Israel (Tel Aviv)
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	6. September 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

Bekannte Probleme

- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem

zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`.

EMRAmazon-Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `RecordServer` Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. `SecretAgent` Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

Version 6.7.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.7.0. Änderungen beziehen sich auf Version 6.6.0.

Erste Version: 15. Juli 2022

Neue Features

- Amazon unterstützt EMR jetzt Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 und Trino 0.378.
- Unterstützt IAM Role- und Lake Formation Formation-basierte Zugriffskontrollen mit EMR Schritten (Spark, Hive) für Amazon EMR auf Clustern. EC2
- Unterstützt Apache-Spark-Datendefinitionsanweisungen auf Apache-Ranger-fähigen Clustern. Dies beinhaltet jetzt die Unterstützung für Trino-Anwendungen, die Apache-Hive-Metadaten auf Apache-Ranger-fähigen Clustern lesen und schreiben. Weitere Informationen finden Sie unter [Aktivieren von Federated Governance mit Trino und Apache Ranger](#) auf Amazon. EMR
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-Amt Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia), Asien-Pazifik (Hyderabad), Naher Osten (UAE),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Europa (Spanien), Europa (Zürich)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 906,0	4,14.322	04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230414.1	4.14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 719.0	4,14.287	10. August 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 606.1	4,14.281	15. Juli 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

Bekannte Probleme

- Wenn EMR Amazon-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, `NoSuchMethodError` tritt ein, weil Amazon ein falsches EMR verwendet. `Hbase.compat.version` EMRAmazon-Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück,

obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. EMRAmazon-Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `RecordServer` Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. `SecretAgent` Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Das wird bei Clustern, die auf Amazon EMR 6.7 oder niedriger ausgeführt werden, GetClusterSessionCredentials API nicht unterstützt.

Version 6.6.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.6.0. Änderungen beziehen sich auf Version 6.5.0.

Erste Version: 09. Mai 2022

Letzte Aktualisierung der Dokumentation: 15. Juni 2022

Neue Features

- Amazon EMR 6.6 unterstützt jetzt Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 und PrestoDB 0.267.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, EMR verwendet Amazon die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 als EMR AMI Standard-A Amazon-Version. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon Linux AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 709.1	4,14.348	23. Juli 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			<p>Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia)</p>

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223,0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906,0		04. Oktober 2023	(US-Ost), China (Peking), China (Ningxia) USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230822.0	4.14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (UAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023_628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230414.1	4.14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230320.0	4.14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 805.0	4,14.287	30. August 2022	us-west-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 719.0	4,14.287	10. August 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 426,0	4,14.281	10. Juni 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 406.1	4,14.275	2. Mai 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

- [Mit Amazon EMR 6.6 und höher werden Anwendungen, die Log4j 1.x und Log4j 2.x verwenden, auf Log4j 1.2.17 \(oder höher\) bzw. Log4j 2.17.1 \(oder höher\) aktualisiert, sodass die zur Behebung der Probleme bereitgestellten Bootstrap-Aktionen nicht erforderlich sind.](#) CVE
- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using](#)

[EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Sizierung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.

Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR reduziert die Cluster-Startzeit für Cluster, die die EMR AMI Standardoption verwenden und nur gängige Anwendungen wie Apache Hadoop, Apache Spark und Apache Hive installieren, um durchschnittlich bis zu 80 Sekunden.

Bekannte Probleme

- Wenn EMR Amazon-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, `NoSuchMethodError` tritt ein, weil Amazon ein falsches EMR verwendet. `Hbase.compat.version` EMRAmazon-Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Connector mit den EMR Versionen 6.6.0, 6.7.0 und 6.8.0 von Spark auf Amazon verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. EMRAmazon-Version 6.9.0 behebt dieses Problem.
- Auf Trino-Clustern mit langer Laufzeit aktiviert Amazon EMR 6.6.0 die Garbage-Collection-Logging-Parameter in der Trino `jvm.config`, um bessere Einblicke in die Garbage-Collection-Logs zu erhalten. Durch diese Änderung werden viele Garbage-Collection-Protokolle an die Datei `launcher.log` (`/var/log/trino/launcher.log`) angehängt. Wenn Sie Trino-Cluster in Amazon EMR 6.6.0 ausführen, kann es vorkommen, dass Knoten aufgrund der angehängten Protokolle nicht mehr genügend Festplattenspeicher haben, nachdem der Cluster einige Tage lang ausgeführt wurde.

Um dieses Problem zu umgehen, führen Sie das folgende Skript als Bootstrap-Aktion aus, um die Garbage-Collection-Logging-Parameter in `jvm.config` zu deaktivieren, während Sie den Cluster für Amazon 6.6.0 erstellen oder klonen. 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"
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- In den EMR Amazon-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei RecordServer Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. SecretAgent Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für Amazon-Versionen EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Version 5.35.0

Dies ist der EMR Versionshinweis zu Amazon-Version 5.35.0.

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.35.0. Änderungen beziehen sich auf Version 5.34.0.

Erste Version: 30. März 2022

Neue Features

- Amazon-Anwendungen der EMR Version 5.35, die Log4j 1.x und Log4j 2.x verwenden, werden auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert und erfordern keine Bootstrap-Aktionen zur Behebung der Probleme in früheren Versionen. CVE Siehe [Ansatz zur Minderung CVE -2021-44228](#).

Änderungen, Verbesserungen und behobene Probleme

Änderungen bei Flink

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> • Aktualisieren Sie die Flink-Version auf 1.14.2. • log4j wurde auf 2.17.1 aktualisiert.

Hadoop-Änderungen

Änderungstyp	Beschreibung
Open-Source-Backports für Hadoop seit 5.34.0 EMR	<ul style="list-style-type: none"> • YARN-10438: Behandle Null in C # () containerId lientRMService getContainerReport • YARN-7266: Threads für den Timeline-Server-Event-Handler sind gesperrt

Änderungstyp	Beschreibung
	<ul style="list-style-type: none">• YARN-10438: ATS 1.5 kann nicht gestartet werden, wenn RollingLevelDb Dateien beschädigt sind oder fehlen• HADOOP-13500: Synchronisieren der Iteration des Objekts mit den Konfigurationseigenschaften• YARN-10651: Ist mit in abgestürzt. CapacityScheduler NPE AbstractYarnScheduler updateNodeResource()• HDFS-12221: Ersetzen Sie Xerces in XmlEditsVisitor• HDFS-16410: Unsicheres Xml-Parsen beim OfflineEditsXmlLoader
Hadoop-Änderungen und Korrekturen	<ul style="list-style-type: none">• Tomcat wird in verwendet und HttpFS wurde auf KMS 8.5.75 aktualisiert• In FileSystemOptimizedCommitter V2 wurde die Erfolgsmarkierung in den commitJob Ausgabepfad geschrieben, der bei der Erstellung des Committers definiert wurde. Da Ausgabepfade commitJob und Ausgabepfade auf Aufgabenebene unterschiedlich sein können, wurde der Pfad korrigiert, sodass er den in den Manifestdateien definierten Pfad verwendet. Bei Hive-Jobs führt dies dazu, dass die Erfolgsmarkierung korrekt geschrieben wird, wenn Operationen wie dynamische Partition oder UNION ALL ausgeführt werden.

Änderungen bei Hive

Änderungstyp	Beschreibung
<p>Hive wurde auf die Open-Source-Version 2.3.9 aktualisiert, einschließlich dieser Korrekturen JIRA</p>	<ul style="list-style-type: none"> • HIVE-17155: findConfFile () in HiveConf .java hat einige Probleme mit dem Conf-Pfad • HIVE-24797: Deaktiviert die Validierung von Standardwerten beim Parsen von Avro-Schemas • HIVE-21563: Verbessern Sie die Leistung von Table#, indem Sie Once deaktivieren getEmptyTable registerAllFunctions • HIVE-18147: Tests können mit java.net fehlschlagen. BindException: Adresse wird bereits verwendet • HIVE-24608: Wechselt im Client für Hive 2.3.x zurück zu get_table HMS • HIVE-21200: Vektorisierung — Die Datumsspalte wirft java.lang aus. UnsupportedOperationException für Parkett • HIVE-19228: Entfernen Sie die Verwendung von commons-httpclient 3.x
<p>Hive Open-Source-Backports seit 5.34.0 EMR</p>	<ul style="list-style-type: none"> • HIVE-19990: Eine Abfrage mit einem Intervall literal in der Join-Bedingung schlägt fehl • HIVE-25824: Aktualisieren Sie Branch-2.3 auf log4j 2.17.0 • TEZ-4062: Die Planung spekulativer Versuche sollte abgebrochen werden, wenn die Aufgabe abgeschlossen ist • TEZ-4108: Während eines spekulativen Ausführungsrennens NullPointerException • TEZ-3918: Die Einstellung tez.task.log.level funktioniert nicht

Änderungstyp	Beschreibung
Hive-Upgrades und -Korrekturen	<ul style="list-style-type: none"> • Aktualisieren Sie die Log4j-Version auf 2.17.1 • Aktualisieren ORC Sie die Version auf 1.4.3 • Der Deadlock aufgrund eines fehlerhaften Thread-Ins wurde behoben ShuffleScheduler
Neue Features	<ul style="list-style-type: none"> • Feature zum Drucken von Hive-Abfragen in AM-Protokollen hinzugefügt. Diese ist standardmäßig deaktiviert. Flag/Konf.: <code>tez.am.emr.print.hive.query.in.log</code> Status (Standard):FALSE.

Oozie-Änderungen

Änderungstyp	Beschreibung
Oozie Open-Source-Backports seit 5.34.0 EMR	<ul style="list-style-type: none"> • OOZIE-3652: Der Oozie-Launcher sollte die Verzeichnisliste erneut versuchen, wenn dies der Fall ist NoSuchFileException

Pig-Änderungen

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> • log4j wurde auf 1.2.17 aktualisiert.

Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.

- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 5.34.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.34.0. Änderungen beziehen sich auf Version 5.33.1.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

Neue Features

- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).
- [Hudi] Verbesserungen zur Vereinfachung der Hudi-Konfiguration. Optimistische Parallelitätssteuerung standardmäßig deaktiviert.

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

- Bisher führte ein manueller Neustart des Resource Managers auf einem Multi-Master-Cluster dazu, dass EMR Amazon-On-Cluster-Daemons wie Zookeeper alle zuvor außer Betrieb genommenen oder verloren gegangenen Knoten in der Zookeeper-Znode-Datei neu laden. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon entfernt EMR jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei und die internen Grenzwerte wurden erhöht.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Zeppelin auf Version 0.10.0 upgegradet.
- Livy Fix – auf 0.7.1 aktualisiert
- Leistungsverbesserung bei Spark — Heterogene Executors werden deaktiviert, wenn bestimmte Spark-Konfigurationswerte in 5.34.0 außer Kraft gesetzt werden. EMR

- Web HDFS - und HttpFS-Server sind standardmäßig deaktiviert. Sie können Web HDFS mithilfe der Hadoop-Konfiguration erneut aktivieren. `dfs.webhdfs.enabled` Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

Bekannte Probleme

- Die Amazon EMR Notebooks-Funktion, die mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den der Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das Notebook mit dem EMR Cluster verbinden. `sudo systemctl start hadoop-httpfs`
- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache Hadoop HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem EMR primären Amazon-Node entweder manuell mithilfe `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen](#) Amazon-Schritt. EMR
- Die Amazon EMR Notebooks-Funktion, die mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den der Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das Notebook mit dem EMR Cluster verbinden. `sudo systemctl start hadoop-httpfs`
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 6.5.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.5.0. Änderungen beziehen sich auf Version 6.4.0.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

Neue Features

- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).
- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Sizierung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.
- Support für das offene Tabellenformat von Apache Iceberg für riesige analytische Datensätze.
- Support für ranger-trino-plugin 2.0.1-amzn-1
- Support für toree 0.5.0

Änderungen, Verbesserungen und behobene Probleme

- Die Release-Version von Amazon EMR 6.5 unterstützt jetzt Apache Iceberg 0.12.0 und bietet Laufzeitverbesserungen mit Amazon EMR Runtime for Apache Spark, Amazon EMR Runtime for Presto und Amazon EMR Runtime for Apache Hive.
- [Apache Iceberg](#) ist ein offenes Tabellenformat für große Datensätze in Amazon S3 und bietet schnelle Abfrageleistung bei großen Tabellen, atomare Commits, gleichzeitige Schreibvorgänge

und SQL -kompatible Tabellenentwicklung. Mit EMR 6.5 können Sie Apache Spark 3.1.2 mit dem Iceberg-Tabellenformat verwenden.

- Apache Hudi 0.9 fügt Spark SQL DDL und Unterstützung hinzu. DML Auf diese Weise können Sie Hudi-Tabellen mit nur Anweisungen erstellen und aktualisieren. SQL Apache Hudi 0.9 beinhaltet auch Leistungsverbesserungen auf der Abfrageseite und auf der Writer-Seite.
- Amazon EMR Runtime for Apache Hive verbessert die Leistung von Apache Hive auf Amazon S3, indem Umbenennungsvorgänge bei Staging-Vorgängen entfernt werden, und verbessert die Leistung von metastore check (MSCK) -Befehlen, die zur Reparatur von Tabellen verwendet werden.

Bekannte Probleme

- Wenn EMR Amazon-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, `NoSuchMethodError` tritt ein, weil Amazon ein falsches EMR verwendet. `Hbase.compat.version` EMRAmazon-Version 6.8.0 behebt dieses Problem.
- Hbase-Bundle-Cluster mit hoher Verfügbarkeit (HA) können nicht mit der Standard-Volumen-Größe und dem Instance-Typ bereitgestellt werden. Dieses Problem lässt sich umgehen, indem Sie die Größe des Root-Volumens erhöhen.
- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-`workflow.xml`-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS E im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie.

Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#\$%&'()*+,-. Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 6.4.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.4.0. Änderungen beziehen sich auf Version 6.3.0.

Erste Version: 20. September 2021

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

Unterstützte Anwendungen

- AWS SDK for Java Version 1.12.31
- CloudWatch Version 2.2.0 senken
- Version 4.16.0 von EMR DynamoDB Connector
- EMRFSAusführung 2.47.0
- Amazon EMR Goodies versie 3.2.0
- Amazon EMR Kinesis Connector, Version 3.5.0
- Amazon EMR Record Server, Version 2.1.0
- Amazon EMR Scripts versie 2.5.0
- Flink, Version 1.13.1
- Ganglia Version 3.7.2
- AWS Glue Hive Metastore Client Version 3.3.0
- Hadoop Version 3.2.1-amzn-4
- HBaseAusführung 2.4.4-amzn-0
- HBase-Operator-Tools 1.1.0
- HCatalogAusführung 3.1.2-amzn-5
- Hive Version 3.1.2-amzn-5
- Hudi Version 0.8.0-amzn-0

- Hue Version 4.9.0
- JDKJava-Version Corretto-8.302.08.1 (Build 1.8.0_302-b08)
- JupyterHub Ausführung 1.4.1
- Livy Version 0.7.1-incubating
- MXNetAusführung 1.8.0
- Oozie-Version 5.2.1
- Phoenix Version 5.1.2
- Pig Version 0.17.0
- Presto Version 0.254.1-amzn-0
- Trino Version 359
- Apache Ranger KMS (transparente Multimaster-Verschlüsselung) Version 2.0.0
- Ranger-Plugins 2.0.1-amzn-0
- Ranger-S3-Plugin 1.2.0
- SageMaker Spark-Version 1.4.1 SDK
- Scala Version 2.12.10 (Offene JDK 64-Bit-Server-VM, Java 1.8.0_282)
- Spark Version 3.1.2-amzn-0
- Spark-Rapids 0.4.1
- Sqoop-Version 1.4.7
- TensorFlow Ausführung 2.4.1
- Tez-Version 0.9.2
- Zeppelin-Version 0.9.0
- Zookeeper Version 3.5.7
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

Neue Features

- [Verwaltete Skalierung] Verwaltete Skalierungsoptimierung mit Spark-Shuffle-Daten — Für EMR Amazon-Versionen 5.34.0 und höher sowie EMR Versionen 6.4.0 und höher ist Managed Scaling jetzt Spark Shuffle-Datenerkennung (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Using EMR Managed Scaling in Amazon EMR](#) im Amazon EMR Management Guide und im [Spark Programming Guide](#).

- Auf Apache Ranger-fähigen EMR Amazon-Clustern können Sie Apache Spark verwenden, SQL um Daten mit, und in die Apache Hive-Metastore-Tabellen einzufügen oder diese zu aktualisieren. INSERT INTO INSERT OVERWRITE ALTER TABLE Bei der Verwendung ALTER TABLE mit Spark SQL muss ein Partitionsspeicherort das untergeordnete Verzeichnis eines Tabellenspeicherorts sein. Amazon unterstützt derzeit EMR nicht das Einfügen von Daten in eine Partition, bei der sich der Speicherort der Partition vom Speicherort der Tabelle unterscheidet.
- Presto SQL wurde [in Trino umbenannt](#).
- Hive: Die Ausführung einfacher SELECT Abfragen mit LIMIT Klausel wird beschleunigt, indem die Abfrageausführung gestoppt wird, sobald die in der LIMIT Klausel angegebene Anzahl von Datensätzen abgerufen wurde. Einfache SELECT Abfragen sind Abfragen ohne ORDER BY/BY-Klausel oder Abfragen, die nicht über eine Reducer-Stufe verfügen. GROUP Beispiel, SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>.

Hudi-Parallelitätskontrolle

- Hudi unterstützt jetzt Optimistic Concurrency Control (OCC), das bei Schreiboperationen genutzt werden kann, z. B. UPSERT INSERT um Änderungen von mehreren Autoren an derselben Hudi-Tabelle zu ermöglichen. Dies ist auf DateiebeneOCC, sodass zwei beliebige Commits (oder Writer) in dieselbe Tabelle schreiben können, sofern ihre Änderungen nicht miteinander in Konflikt stehen. Weitere Informationen finden Sie unter [Hudi-Parallelitätskontrolle](#).
- Amazon EMR Amazon-Clustern ist Zookeeper installiert, der als Sperranbieter für verwendet werden kann. OCC Um die Verwendung dieser Funktion zu vereinfachen, sind in EMR Amazon-Clustern die folgenden Eigenschaften vorkonfiguriert:

```
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
```

Zur Aktivierung OCC müssen Sie die folgenden Eigenschaften entweder mit ihren Hudi-Joboptionen oder auf Cluster-Ebene mithilfe der Amazon-Konfigurationen konfigurieren: EMR API

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table Name is a good option)
```

Hudi Monitoring: CloudWatch Amazon-Integration zur Berichterstattung über Hudi-Metriken

- Amazon EMR unterstützt die Veröffentlichung von Hudi Metrics auf Amazon CloudWatch. Es wird aktiviert, indem die folgenden erforderlichen Konfigurationen festgelegt werden:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Im Folgenden finden Sie optionale Hudi-Konfigurationen, die Sie ändern können:

Einstellung	Beschreibung	Wert
hoodie.metrics.cloudwatch.report.period.seconds	Häufigkeit (in Sekunden), mit der Kennzahlen an Amazon gemeldet werden CloudWatch	Der Standardwert ist 60s, was für die von Amazon angebotene Standardauflösung von einer Minute in Ordnung ist CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Präfix, das jedem Metriknamen hinzugefügt werden soll	Der Standardwert ist leer (kein Präfix)
hoodie.metrics.cloudwatch.namespace	CloudWatch Amazon-Namespace, unter dem Metriken veröffentlicht werden	Der Standardwert ist Hudi
hoodie.metrics.cloudwatch.maxDatumsPerAnfrage	Maximale Anzahl von Daten, die in einer Anfrage an Amazon enthalten sein können CloudWatch	Der Standardwert ist 20, was dem CloudWatch Amazon-Standard entspricht

Unterstützung und Verbesserungen von Amazon EMR Hudi-Konfigurationen

- Kunden können jetzt die Funktion „EMRKonfigurationen API und Neukonfiguration“ nutzen, um Hudi-Konfigurationen auf Cluster-Ebene zu konfigurieren. Eine neue dateibasierte Konfigurationsunterstützung wurde über `/etc/hudi/conf/hudi-defaults.conf` eingeführt, ähnlich wie bei anderen Anwendungen wie Spark, Hive usw. EMRkonfiguriert einige Standardeinstellungen, um die Benutzererfahrung zu verbessern:

- `hoodie.datasource.hive_sync.jdbcurl` ist für den Cluster-Hive-Server konfiguriert URL und muss nicht mehr angegeben werden. Dies ist besonders nützlich, wenn Sie einen Job im Spark-Cluster-Modus ausführen, wo Sie zuvor die EMR Amazon-Master-IP angeben mussten.
- HBase spezifische Konfigurationen, die für die Verwendung des HBase Index mit Hudi nützlich sind.
- Spezifische Konfiguration des Zookeeper-Sperranbieters, wie unter Parallelitätskontrolle beschrieben, was die Verwendung von Optimistic Concurrency Control () erleichtert. OCC
- Zusätzliche Änderungen wurden eingeführt, um die Anzahl der Konfigurationen zu reduzieren, die Sie bestehen müssen, und um nach Möglichkeit automatische Schlüsse zu ziehen:
 - Das `partitionBy` -Schlüsselwort kann verwendet werden, um die Partitionsspalte zu spezifizieren.
 - Bei der Aktivierung von Hive Sync ist es nicht mehr erforderlich, den Vorgang `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY` zu bestehen. Diese Werte können aus dem Hudi-Tabellennamen und dem Partitionsfeld abgeleitet werden.
 - `KEYGENERATOR_CLASS_OPT_KEY` ist nicht zwingend erforderlich und kann aus einfacheren Fällen von `SimpleKeyGenerator` und `ComplexKeyGenerator` abgeleitet werden.

Vorbehalte von Hudi

- Hudi unterstützt keine vektorisierte Ausführung in Hive für Merge on Read (MoR)- und Bootstrap-Tabellen. So schlägt beispielsweise `count(*)` mit der Hudi-Echtzeittabelle fehl, wenn `hive.vectorized.execution.enabled` auf „wahr“ gesetzt ist. Um das Problem zu umgehen, können Sie das vektorisierte Lesen deaktivieren, indem Sie `hive.vectorized.execution.enabled` auf `false` setzen.
- Die Multi-Writer-Unterstützung ist nicht mit dem Hudi-Bootstrap-Feature kompatibel.
- Flink Streamer und Flink SQL sind experimentelle Funktionen in dieser Version. Diese Features werden nicht zur Verwendung in Produktionsbereitstellungen empfohlen.

Änderungen, Verbesserungen und behobene Probleme

Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.

- Bisher führte ein manueller Neustart des Resource Managers auf einem Multi-Master-Cluster dazu, dass EMR Amazon-On-Cluster-Daemons wie Zookeeper alle zuvor außer Betrieb genommenen oder verloren gegangenen Knoten in der Zookeeper-Znode-Datei neu laden. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon entfernt EMR jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei und die internen Grenzwerte wurden erhöht.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Konfiguration eines Clusters zur Behebung von Leistungsproblemen mit Apache YARN Timeline Server Version 1 und 1.5

Apache YARN Timeline Server Version 1 und 1.5 können bei sehr aktiven, großen EMR Clustern zu Leistungsproblemen führen, insbesondere bei `yarn.resourcemanager.system-metrics-publisher.enabled=true`, was die Standardeinstellung in Amazon ist EMR. Ein YARN Open-Source-Timeline Server v2 löst das Leistungsproblem im Zusammenhang mit der Skalierbarkeit von YARN Timeline Server.

Weitere Lösungen für dieses Problem umfassen:

- Konfiguration von `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- Aktivieren des Fixes für dieses Problem beim Erstellen eines Clusters, wie unten beschrieben.

Die folgenden EMR Amazon-Versionen enthalten eine Lösung für dieses Leistungsproblem mit dem YARN 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

Um den Fix für eine der oben angegebenen EMR Amazon-Versionen zu aktivieren, legen Sie diese Eigenschaften `true` in einer JSON Konfigurationsdatei fest, die mit dem [aws emr create-cluster Befehlsparameter](#) übergeben wird: `--configurations file://./configurations.json`. Oder aktivieren Sie den Fix über die Benutzeroberfläche der [Rekonfigurationskonsole](#).

Beispiel für den Inhalt der Datei `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": []
}
```

- Web HDFS - und HttpFS-Server sind standardmäßig deaktiviert. Sie können Web HDFS mithilfe der Hadoop-Konfiguration erneut aktivieren. `dfs.webhdfs.enabled` Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

- HTTPS ist jetzt standardmäßig für Amazon Linux-Repositorys aktiviert. Wenn Sie eine Amazon S3 VPC S3-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon Linux-Bucket ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (`$region` ersetzen Sie ihn durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Verwendung HTTPS beim Herstellen einer Verbindung zu Paket-Repositorys](#).
- Hive: Die Leistung bei Schreibabfragen wurde verbessert, indem die Verwendung eines Scratch-Verzeichnisses HDFS für den letzten Job aktiviert wurde. Die temporären Daten für den endgültigen Job werden HDFS nicht in Amazon S3 geschrieben, und die Leistung wird verbessert, da die Daten vom HDFS Speicherort der endgültigen Tabelle (Amazon S3) und nicht zwischen Amazon S3 S3-Geräten verschoben werden.
- Hive: Verbesserung der Kompilierungszeit von Abfragen um das 2,5-fache mit Glue Metastore Partition Pruning.
- Wenn Built-In von Hive an den Hive Metastore Server übergeben UDFs werden, wird standardmäßig nur ein Teil der UDFs Built-In an den Glue Metastore übergeben, da Glue nur begrenzte Ausdrucksoperatoren unterstützt. Wenn Sie `hive.glue.partition.pruning.client=true` festlegen, erfolgt das gesamte Partitionsbereinigen auf der Clientseite. Wenn Sie `hive.glue.partition.pruning.server=true` festlegen, erfolgt das gesamte Bereinigen von Partitionen auf der Serverseite.

Bekannte Probleme

- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache Hadoop HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem EMR primären Amazon-Node entweder manuell mithilfe `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen](#) Amazon-Schritt. EMR
- Die Amazon EMR Notebooks-Funktion, die mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den der Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das Notebook mit dem EMR Cluster verbinden. `sudo systemctl start hadoop-httpfs`
- In EMR Amazon-Version 6.4.0 unterstützt Phoenix die Komponente Phoenix Connectors nicht.

- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-workflow.xml-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS E im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 5.32.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.32.0. Änderungen beziehen sich auf Version 5.31.0.

Erste Version: 8. Januar 2021

Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.14.0 aktualisiert
- Amazon SageMaker Spark wurde SDK auf Version 1.4.1 aktualisiert
- Auf Version AWS SDK for Java 1.11.890 aktualisiert
- EMRDynamoDB Connector Version 4.16.0 wurde aktualisiert

- Auf Version 2.45.0 aktualisiert EMRFS
- EMRLog Analytics-Metriken wurden auf Version 1.18.0 aktualisiert
- EMR MetricsAndEventsApiGateway Der Client wurde auf Version 1.5.0 aktualisiert
- EMRRecord Server wurde auf Version 1.8.0 aktualisiert
- EMRS3 Dist CP wurde auf Version 2.17.0 aktualisiert
- EMRSecret Agent wurde auf Version 1.7.0 aktualisiert
- Flink auf Version 1.11.2 aktualisiert
- Hadoop wurde auf Version 2.10.1-amzn-0 aktualisiert
- Upgrade von Hive auf Version 2.3.7-amzn-3
- Hue auf Version 4.8.0 aktualisiert
- Mxnet wurde auf Version 1.7.0 aktualisiert
- OpenCV wurde auf Version 4.4.0 aktualisiert
- Presto auf Version 0.240.1-amzn-0 aktualisiert
- Spark auf Version 2.4.7-amzn-0 aktualisiert
- Auf Version TensorFlow 2.3.1 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.

- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Aktualisierte Komponentenversionen.
- Eine Liste der Komponentenversionen finden Sie unter [About Amazon EMR Releases](#) in diesem Handbuch.

Neue Features

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Sizierung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.
- Unterstützungsstatus von Instance Metadata Service (IMDS) V2: Die Komponenten Amazon EMR 5.23.1, 5.27.1 und 5.32 oder höher werden für alle Aufrufe verwendet. IMDSv2 IMDS Für IMDS Aufrufe in Ihrem Anwendungscode können Sie sowohl als auch IMDSv1 verwenden oder die Option so konfigurieren `IMDSv2`, dass sie nur für zusätzliche Sicherheit IMDS verwendet wird. IMDSv2 Bei anderen EMR 5.x-Versionen IMDSv1 führt die Deaktivierung zu einem Fehler beim Starten des Clusters.
- Ab Amazon EMR 5.32.0 können Sie einen Cluster starten, der nativ in Apache Ranger integriert ist. Apache Ranger ist ein Open-Source-Framework zur Aktivierung, Überwachung und Verwaltung einer umfassenden Datensicherheit auf der gesamten Hadoop-Plattform. Weitere Informationen finden Sie unter [Apache Ranger](#). Dank der nativen Integration können Sie Ihren eigenen Apache Ranger verwenden, um eine detaillierte Datenzugriffskontrolle auf Amazon durchzusetzen. EMR

Weitere Informationen finden Sie unter [Integrieren von Amazon EMR mit Apache Ranger](#) im EMRAmazon-Versionshandbuch.

- Amazon EMR Release 5.32.0 unterstützt Amazon EMR on. EKS Weitere Informationen zu den ersten Schritten mit EMR on EKS finden Sie unter [Was ist Amazon EMR auf EKS](#).
- Amazon EMR Release 5.32.0 unterstützt Amazon EMR Studio (Preview). Weitere Informationen zu den ersten Schritten mit EMR Studio finden Sie unter [Amazon EMR Studio \(Vorversion\)](#).
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, EMR hat Amazon EMR verwaltete Standardrichtlinien mit Geltungsbereich v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [EMRVon Amazon verwaltete Richtlinien](#).

Bekanntete Probleme

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Websites wie SparkUIs, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMRAmazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von ofAmazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-StandardEinstellung eine Ulimit-StandardEinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

⚠ Important

EMR Cluster, auf denen Amazon Linux oder Amazon Linux 2 Amazon Machine Images (AMIs) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen EC2 Amazon-Instances, auf denen das standardmäßige Amazon Linux ausgeführt wird. Wenn neue Amazon Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel und CUDA Updates) NVIDIA, verfügbar werden, nachdem eine EMR Amazon-Version verfügbar wird, laden EMR Cluster-Instances, die standardmäßig ausgeführt werden, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon so anpassen EMR AMI](#), dass es [das neueste Amazon Linux verwendet AMI](#).

- Die Konsolenunterstützung zur Erstellung einer Sicherheitskonfiguration, die die AWS Ranger-Integrationsoption spezifiziert, wird derzeit in der GovCloud Region nicht unterstützt. Die Sicherheitskonfiguration kann mit dem CLI vorgenommen werden. Weitere Informationen finden Sie unter [EMR Sicherheitskonfiguration erstellen](#) im Amazon EMR Management Guide.
- Wenn AtRestEncryption die HDFS Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
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
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.

- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U +0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 6.2.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.2.0.

Änderungen beziehen sich auf Version 6.1.0.

Veröffentlichungsdatum: 9. Dezember 2020

Letzte Aktualisierung: 4. Oktober 2021

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.828
- emr-record-server Ausführung 1.7.0
- Flink-Version 1.11.2
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBaseAusführung 2.2.6-amzn-0
- HBase-Operator-Tools 1.0.0
- HCatalogAusführung 3.1.2-amzn-0
- Hive Version 3.1.2-amzn-3
- Hudi Version 0.6.0-amzn-1
- Hue Version 4.8.0
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0

- MXNetAusführung 1.7.0
- Oozie-Version 5.2.0
- Phoenix Version 5.0.0
- Pig Version 0.17.0
- Presto Version 0.238.3-amzn-1
- Presto, Version 3.4.3 SQL
- Spark Version 3.0.1-amzn-0
- Spark-Rapids 0.2.0
- TensorFlow Ausführung 2.3.1
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

Neue Features

- HBase: Die Umbenennung in der Commit-Phase wurde entfernt und persistentes HFile Tracking hinzugefügt. Weitere Informationen finden Sie unter [Persistent HFile Tracking](#) im Amazon EMR Release Guide.
- HBase: Backported [Erstellen Sie eine Konfiguration, die das Zwischenspeichern von Blöcken bei der Komprimierung erzwingt](#).
- PrestoDB: Verbesserungen beim dynamischen Partitionsbereinigen. Die regelbasierte Join Reorder funktioniert mit nicht partitionierten Daten.
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, EMR hat Amazon EMR verwaltete Standardrichtlinien mit Geltungsbereich v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [EMRVon Amazon verwaltete Richtlinien](#).
- Support-Status für Instance Metadata Service (IMDS) V2: Für Amazon EMR 6.2 oder höher werden EMR Amazon-Komponenten IMDSv2 für alle IMDS Aufrufe verwendet. Für IMDS Aufrufe in Ihrem Anwendungscode können Sie IMDSv1 sowohl als auch verwenden oder das so konfigurierenIMDSv2, IMDS dass es nur aus IMDSv2 Sicherheitsgründen verwendet wird. Wenn Sie die Option IMDSv1 in früheren Versionen von Amazon EMR 6.x deaktivieren, führt dies zu einem Fehler beim Starten des Clusters.

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Spark: Leistungsverbesserungen in der Spark-Laufzeit.

Bekannte Probleme

- Amazon EMR 6.2 hat falsche Berechtigungen für die Datei `libinstance-controller-java /etc/cron.d/` in 6.2.0 festgelegt. EMR Die Berechtigungen für die Datei sind 645 (-rw-r--r-x), obwohl sie 644 (-rw-r--r--) sein sollten. Aus diesem Grund protokolliert Amazon EMR Version 6.2 keine Instance-State-Logs und das Verzeichnis `/emr/instance-logs` ist leer. Dieses Problem wurde in Amazon EMR 6.3.0 und höher behoben.

Zur Umgehung dieses Problems führen Sie das folgende Skript als Bootstrap-Aktion beim Clusterstart aus.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Für private Subnetz-Cluster von Amazon EMR 6.2.0 und 6.3.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Websites wie SparkUIs, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den `httpd`-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- In Amazon EMR 6.2.0 gibt es ein Problem, bei dem `httpd` kontinuierlich ausfällt, wodurch Ganglia nicht verfügbar ist. Sie erhalten die Fehlermeldung „Es kann keine Verbindung zum Server hergestellt werden“. Um einen Cluster zu reparieren, der bereits mit diesem Problem läuft, gehen Sie SSH zum primären Clusterknoten und fügen Sie die Zeile `Listen 80` zu der Datei `httpd.conf` hinzu, die sich unter befindet. `/etc/httpd/conf/httpd.conf` Dieses Problem wurde in Amazon EMR 6.3.0 behoben.
- HTTPDSchlägt auf EMR 6.2.0-Clustern fehl, wenn Sie eine Sicherheitskonfiguration verwenden. Dadurch ist die Benutzeroberfläche der Ganglia-Webanwendung nicht verfügbar. Um auf die Benutzeroberfläche der Ganglia-Webanwendung zuzugreifen, fügen Sie `Listen 80` der `/etc/httpd/conf/httpd.conf`-Datei auf dem Primärknoten Ihres Clusters etwas hinzu. Informationen zum Herstellen einer Verbindung mit Ihrem Cluster finden Sie unter [Herstellen einer Verbindung zum Primärknoten mithilfe von SSH](#).

EMRNotebooks können auch keine Verbindung zu EMR 6.2.0-Clustern herstellen, wenn Sie eine Sicherheitskonfiguration verwenden. Das Notebook kann keine Kernel auflisten und Spark-Aufträge

nicht weiterleiten. Wir empfehlen, EMR Notebooks EMR stattdessen mit einer anderen Version von Amazon zu verwenden.

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein `ulimit` über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein `ulimit` festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich entscheidend auf alle Hudi-Einfüge-, Upsert- und Löschvorgänge auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi zu erhalten. RPM

• **⚠ Important**

EMRCluster, auf denen Amazon Linux oder Amazon Linux 2 Amazon Machine Images (AMIs) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen EC2 Amazon-Instances, auf denen das standardmäßige Amazon Linux ausgeführt wird. AMI. Wenn neue Amazon Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel und CUDA Updates) NVIDIA, verfügbar werden, nachdem eine EMR Amazon-Version verfügbar wird, laden EMR Cluster-Instances, die standardmäßig ausgeführt werden, diese Updates AMI

nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon so anpassen EMR AMI](#), dass es [das neueste Amazon Linux verwendet AMI](#).

- Amazon EMR 6.2.0 Maven-Artefakte werden nicht veröffentlicht. Sie werden mit einer future Version von Amazon veröffentlichtEMR.
- Die persistente HFile Nachverfolgung mithilfe der HBase Storefile-Systemtabelle unterstützt die Funktion zur HBase Regionsreplikation nicht. Weitere Informationen zur HBase Regionsreplikation finden Sie unter [Timeline-consistent](#) High Available Reads.
- Unterschiede zwischen den Bucketing-Versionen von Amazon EMR EMR 6.x und 5.x Hive

EMR5.x verwendet OOS Apache Hive 2, während in EMR 6.x Apache Hive 3 verwendet wird. OOS Die Open-Source-Version Hive2 verwendet Bucketing Version 1, während die Open-Source-Version Hive3 Bucketing Version 2 verwendet. Dieser Unterschied zwischen Hive 2 (EMR5.x) und Hive 3 (EMR6.x) bedeutet, dass Hive Bucketing-Hashing anders funktioniert. Sehen Sie sich das folgende Beispiel an.

Die folgende Tabelle ist ein Beispiel, das in 6.x bzw. 5.x erstellt wurde. EMR EMR

```
-- 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/';
```

Dieselben Daten werden sowohl in EMR 6.x als auch in 5.x eingefügt. EMR

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Die Überprüfung des S3-Speicherorts zeigt, dass der Name der Bucketing-Datei unterschiedlich ist, da sich die Hashing-Funktion zwischen EMR 6.x (Hive 3) und 5.x (Hive 2) unterscheidet. EMR

```
[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
```

Sie können den Versionsunterschied auch erkennen, indem Sie den folgenden Befehl in Hive in 6.x ausführen. CLI EMR Beachten Sie, dass die Bucketing-Version 2 zurückgegeben wird.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 5.31.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.31.0. Änderungen beziehen sich auf Version 5.30.1.

Erste Version: 09. Oktober 2020

Letzte Aktualisierung: 15. Oktober 2020

Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.13.0 aktualisiert
- Amazon SageMaker Spark wurde SDK auf Version 1.4.0 aktualisiert
- Der Amazon-Kinesis-Konnektor wurde auf Version 3.5.9 aktualisiert
- Auf Version AWS SDK for Java 1.11.852 aktualisiert
- BigTop-Tomcat wurde auf Version 8.5.56 aktualisiert

- EMRFS wurde auf Version 2.43.0 aktualisiert
- EMR MetricsAndEventsApiGateway Der Client wurde auf Version 1.4.0 aktualisiert
- EMRS3 Dist CP wurde auf Version 2.15.0 aktualisiert
- EMRS3 Select wurde auf Version 1.6.0 aktualisiert
- Flink wurde auf Version 1.11.0 aktualisiert
- Hadoop wurde auf Version 2.10.0 aktualisiert
- Hive wurde auf Version 2.3.7 aktualisiert
- Hudi wurde auf Version 0.6.0 aktualisiert
- Hue wurde auf Version 4.7.1 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Mxnet wurde auf Version 1.6.0 aktualisiert
- OpenCV wurde auf Version 4.3.0 aktualisiert
- Presto wurde auf Version 0.238.3 aktualisiert
- Auf Version TensorFlow 2.1.0 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.

- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- [Hive-Spaltenstatistiken](#) werden für EMR Amazon-Versionen 5.31.0 und höher unterstützt.
- Aktualisierte Komponentenversionen.
- EMRFSS3EC V2-Unterstützung in Amazon EMR 5.31.0. In den SDK S3-Java-Versionen 1.11.837 und höher wurde der Verschlüsselungsclient Version 2 (S3EC V2) mit verschiedenen Sicherheitsverbesserungen eingeführt. Weitere Informationen finden Sie hier:
 - S3 Blogbeitrag: [Updates für den Amazon S3 Encryption Client](#).
 - AWS SDK for Java Entwicklerhandbuch: [Migrieren Sie Verschlüsselungs- und Entschlüsselungsclients auf V2](#).
 - EMRVerwaltungshandbuch: [Clientseitige Amazon S3 S3-Verschlüsselung](#).

Der Encryption Client V1 ist aus SDK Gründen der Abwärtskompatibilität weiterhin in der verfügbar.

Neue Features

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMRAmazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von ofAmazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit

einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine Ulimit-Standard-Einstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- Mit Amazon EMR 5.31.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration ermöglicht eine detaillierte Datenfilterung auf Spaltenebene für Datenbanken und Tabellen im Glue-Datenkatalog. AWS ermöglicht auch föderiertes Single Sign-On für EMR Notebooks oder Apache Zeppelin von einem Unternehmensidentitätssystem aus. Weitere Informationen finden Sie unter [Integration von Amazon EMR mit AWS Lake Formation](#) im Amazon EMR Management Guide.

Amazon EMR with Lake Formation ist derzeit in 16 AWS Regionen verfügbar: USA Ost (Ohio und Nord-Virginia), USA West (Nordkalifornien und Oregon), Asien-Pazifik (Mumbai, Seoul, Singapur, Sydney und Tokio), Kanada (Zentral), Europa (Frankfurt, Irland, London, Paris und Stockholm), Südamerika (São Paulo).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn AtRestEncryption die HDFS Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
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
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie.

Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#\$%&'()*+,-.
Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 6.1.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 6.1.0.
Änderungen beziehen sich auf Version 6.0.0.

Erste Version: 4. September 2020

Letzte Aktualisierung: 15. Oktober 2020

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.828
- Flink-Version 1.11.0
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBaseAusführung 2.2.5
- HBase-Operator-Tools 1.0.0
- HCatalogAusführung 3.1.2-amzn-0
- Hive-Version 3.1.2-amzn-1
- Hudi-Version 0.5.2-incubating
- Hue-Version 4.7.1
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0
- MXNetAusführung 1.6.0
- Oozie-Version 5.2.0
- Phoenix Version 5.0.0
- Presto-Version 0.232
- Presto, Version 3.38 SQL

- Spark-Version 3.0.0-amzn-0
- TensorFlow Ausführung 2.1.0
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

Neue Features

- ARM Instance-Typen werden ab EMR Amazon-Version 5.30.0 und EMR Amazon-Version 6.1.0 unterstützt.
- Allzweck-Instance-Typen von M6g werden ab den EMR Amazon-Versionen 6.1.0 und 5.30.0 unterstützt. Weitere Informationen finden Sie unter [Unterstützte Instance-Typen](#) im Amazon EMR Management Guide.
- Die EC2 Platzierungsgruppenfunktion wird ab EMR Amazon-Version 5.23.0 als Option für Cluster mit mehreren primären Knoten unterstützt. Derzeit werden nur Primärknoten-Typen von des Platzierungsgruppenfeatures unterstützt, und die SPREAD-Strategie wird auf diese Primärknoten angewendet. Bei dieser SPREAD-Strategie wird eine kleine Gruppe von Instances auf separater zugrundeliegender Hardware platziert, um den Verlust mehrerer Primärknoten im Falle eines Hardwarefehlers zu verhindern. Weitere Informationen finden Sie unter [EMR Integration mit EC2 Placement Group](#) im Amazon EMR Management Guide.
- Verwaltete Skalierung — Mit EMR Amazon-Version 6.1.0 können Sie Amazon EMR Managed Scaling aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster je nach Arbeitslast automatisch zu erhöhen oder zu verringern. Amazon wertet EMR kontinuierlich Cluster-Metriken aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster im Hinblick auf Kosten und Geschwindigkeit optimieren. Managed Scaling ist auch auf EMR Amazon-Version 5.30.0 und höher verfügbar, außer 6.0.0. Weitere Informationen finden Sie unter [Scaling Cluster Resources](#) im Amazon EMR Management Guide.
- Presto SQL Version 338 wird mit EMR 6.1.0 unterstützt. Weitere Informationen finden Sie unter [Presto](#).
 - Presto SQL wird nur in EMR 6.1.0 und späteren Versionen unterstützt, nicht in 6.0.0 oder 5.x.
EMR EMR
 - Der Anwendungsname Presto wird weiterhin verwendet, um PrestoDB auf Clustern zu installieren. Verwenden Sie den Anwendungsnamen, um Presto SQL auf Clustern zu installieren.
PrestoSQL

- Sie können entweder PrestoDB oder Presto installieren, aber Sie können nicht beide auf einem einzigen Cluster installieren. Wenn beim Versuch, einen Cluster zu erstellen, sowohl PrestoDB als auch Presto angegeben werden, tritt ein Validierungsfehler auf und die Anfrage zur Clustererstellung schlägt fehl.
- Presto SQL wird sowohl auf Single-Master- als auch auf Multi-Master-Clustern unterstützt. Auf Multi-Master-Clustern ist ein externer Hive-Metastore erforderlich, um Presto oder SQL PrestoDB auszuführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR Cluster mit mehreren Primärknoten](#).
- Unterstützung der auto Authentifizierung auf Apache Hadoop und Apache Spark mit Docker: Spark-Benutzer können Docker-Images von Docker Hub und Amazon Elastic Container Registry (Amazon ECR) verwenden, um Umgebungs- und Bibliotheksabhängigkeiten zu definieren.

[Konfigurieren Sie Docker](#) und [führen Sie Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x](#) aus.

- EMR unterstützt Apache ACID Hive-Transaktionen: Amazon EMR 6.1.0 bietet Unterstützung für ACID Hive-Transaktionen, sodass es den ACID Eigenschaften einer Datenbank entspricht. Mit diesem Feature können Sie INSERT, UPDATE, DELETE, - und MERGE-Operationen in Hive-Tabellen mit Daten in Amazon Simple Storage Service (Amazon S3) ausführen. Dies ist eine wichtige Funktion für Anwendungsfälle wie Streaming-Erfassung, Neuformulierung von Daten, Verwendung MERGE von Massenaktualisierungen und langsam ändernde Dimensionen. Weitere Informationen, einschließlich Konfigurationsbeispielen und Anwendungsfällen, finden Sie unter [Amazon EMR unterstützt Apache ACID Hive-Transaktionen](#).

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.

- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Apache Flink wird auf 6.0.0 nicht unterstützt, aber es wird auf 6.1.0 mit Flink 1.11.0 unterstützt. EMR Dies ist die erste Version von Flink, die Hadoop 3 offiziell unterstützt. Siehe Ankündigung der Veröffentlichung von [Apache Flink 1.11.0](#).
- Ganglia wurde aus den Standard-6.1.0-Paketpaketen entfernt. EMR

Bekannte Probleme

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von ofAmazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standardeinstellung eine Ulimit-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit

von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
```

```
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich entscheidend auf alle Hudi-Einfüge-, Upsert- und Löschvorgänge auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi zu erhalten. RPM

- Wenn Sie eine benutzerdefinierte Garbage-Collection-Konfiguration mit `spark.driver.extraJavaOptions` und festlegen, führt dies `dazuspark.executor.extraJavaOptions`, dass der Start von Treiber/Executor mit 6.1 aufgrund einer widersprüchlichen Garbage-Collection-Konfiguration fehlschlägt. EMR In EMR Version 6.1.0 sollten Sie stattdessen eine benutzerdefinierte Spark-Garbage-Collection-Konfiguration für Treiber und Executors mit den Eigenschaften und angeben. `spark.driver.defaultJavaOptions` `spark.executor.defaultJavaOptions` Weitere Informationen finden Sie unter [Apache Spark Runtime Environment](#) und [Konfiguration von Spark Garbage Collection auf Amazon EMR 6.1.0](#).
- Die Verwendung von Pig mit Oozie (und innerhalb von Hue, da Hue Oozie-Aktionen zur Ausführung von Pig-Skripten verwendet) generiert den Fehler, dass eine native LZ0-Bibliothek nicht geladen werden kann. Diese Fehlermeldung ist informativ und verhindert nicht, dass Pig ausgeführt wird.
- Hudi-Parallelitätsunterstützung: Derzeit unterstützt Hudi keine gleichzeitigen Schreibvorgänge in eine einzelne Hudi-Tabelle. Darüber hinaus macht Hudi alle Änderungen rückgängig, die von in Bearbeitung befindlichen Autoren vorgenommen wurden, bevor ein neuer Writer beginnen kann. Gleichzeitige Schreibvorgänge können diesen Mechanismus stören und zu Wettlaufbedingungen führen, was zu Datenbeschädigungen führen kann. Sie sollten sicherstellen, dass im Rahmen Ihres Datenverarbeitungs-Workflows immer nur ein einziger Hudi-Writer mit einer Hudi-Tabelle arbeitet. Hudi unterstützt mehrere gleichzeitige Lesegeräte, die mit derselben Hudi-Tabelle arbeiten.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-

Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- In Amazon EMR 6.1.0 gibt es ein Problem, das Cluster betrifft, auf denen Presto ausgeführt wird. Nach einem längeren Zeitraum (Tage) kann der Cluster Fehler wie „su: failed to execute /bin/bash: Resource temporarily unavailable“ oder „Shell-Anfrage auf Kanal 0 fehlgeschlagen“ ausgeben. Dieses Problem wird durch einen internen EMR Amazon-Prozess (InstanceController) verursacht, der zu viele Light-Weight-Prozesse (LWP) erzeugt, was letztendlich dazu führt, dass der Hadoop-Benutzer sein Nproc-Limit überschreitet. Dadurch wird verhindert, dass der Benutzer zusätzliche Prozesse öffnet. Die Lösung für dieses Problem ist ein Upgrade auf 6.2.0. EMR

Version 6.0.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 6.0.0.

Erste Version: 10. März 2020

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.711
- Ganglia Version 3.7.2
- Hadoop Version 3.2.1
- HBaseAusführung 2.2.3
- HCatalogAusführung 3.1.2
- Hive Version 3.1.2
- Hudi Version 0.5.0-incubating
- Hue Version 4.4.0
- JupyterHub Ausführung 1.0.0
- Livy Version 0.6.0
- MXNetAusführung 1.5.1
- Oozie Version 5.1.0
- Phoenix Version 5.0.0
- Presto Version 0.230
- Spark Version 2.4.4
- TensorFlow Ausführung 1.14.0
- Zeppelin versie 0.9.0- SNAPSHOT
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

Note

Flink, Sqoop, Pig und Mahout sind in der EMR Amazon-Version 6.0.0 nicht verfügbar.

Neue Features

- YARN Docker Runtime Support — YARN Anwendungen wie Spark-Jobs können jetzt im Kontext eines Docker-Containers ausgeführt werden. Auf diese Weise können Sie auf einfache Weise Abhängigkeiten in einem Docker-Image definieren, ohne benutzerdefinierte Bibliotheken auf Ihrem

EMR Amazon-Cluster installieren zu müssen. Weitere Informationen finden [Sie unter Docker-Integration konfigurieren](#) und [Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.0.0 ausführen](#).

- LLAPHive-Unterstützung — Hive unterstützt jetzt den LLAP Ausführungsmodus für eine verbesserte Abfrageleistung. Weitere Informationen finden Sie unter Hive [verwenden](#). LLAP

Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon EMR Scaling, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder wenn es zu Anwendungsausfällen kommt.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn EMR Amazon-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen von YARN Knotenstatus und HDFS Knotenstatus. Dies geschah, weil On-Cluster-Daemons nicht in der Lage waren, die Gesundheitsstatusdaten eines Knotens an interne Amazon-Komponenten zu übermitteln. EMR
- Die EMR Cluster-Daemons wurden verbessert, um den Knotenstatus bei der Wiederverwendung von IP-Adressen korrekt nachzuverfolgen, um die Zuverlässigkeit bei Skalierungsvorgängen zu erhöhen.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Jobfehler aufgrund eines Fehlers bei der YARN Außerbetriebnahme auftraten, wenn der Cluster versuchte, nach oben oder unten zu skalieren.
- Das Problem mit Schritt- oder Jobfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den EMR Amazon-On-Cluster-Daemons und/ immer konsistent sind. YARN HDFS
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für EMR Amazon-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der EMR Amazon-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation HDFS YARN mit/auf dem primären Knoten erforderlich ist.
- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen

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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.

- Amazon Linux
 - Amazon Linux 2 ist das Betriebssystem für die EMR 6.x-Release-Serie.
 - `systemd` wird anstelle von `in` Amazon Linux 1 für das Servicemanagement `upstart` verwendet.
- Java-Entwicklungskit (JDK)
 - Corretto JDK 8 ist der Standard JDK für die EMR 6.x-Release-Serie.
- Scala
 - Scala 2.12 wird mit Apache Spark und Apache Livy verwendet.
- Python 3
 - Python 3 ist jetzt die Standardversion von Python in EMR.
- YARN Knotenbeschriftungen
 - Ab der Amazon EMR 6.x-Release-Serie ist die Funktion YARN Node Labels standardmäßig deaktiviert. Die Anwendungs-Master-Prozesse können standardmäßig sowohl auf Kern- als auch auf Aufgabenknoten ausgeführt werden. Sie können die Funktion „YARN Node Labels“ aktivieren, indem Sie die folgenden Eigenschaften konfigurieren: `yarn.node-labels.enabled` und `yarn.node-labels.am.default-node-label-expression`. Weitere Informationen finden Sie unter [Grundlegendes zu Primär-, Kern- und Aufgabenknoten](#).

Bekannte Probleme

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standard-Einstellung eine `ulimit`-Standard-Einstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt

Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- Die interaktive Spark-Shell PySpark, einschließlich SparkR und Spark-Shell, unterstützt die Verwendung von Docker mit zusätzlichen Bibliotheken nicht.
- Um Python 3 mit Amazon EMR Version 6.0.0 zu verwenden, müssen Sie Folgendes PATH hinzufügen `yarn.nodemanager.env-whitelist`.
- Die Funktionen Live Long und Process (LLAP) werden nicht unterstützt, wenn Sie den AWS Glue-Datenkatalog als Metastore für Hive verwenden.
- Wenn Sie Amazon EMR 6.0.0 mit Spark- und Docker-Integration verwenden, müssen Sie die Instances in Ihrem Cluster mit demselben Instance-Typ und derselben Anzahl von EBS Volumes konfigurieren, um Fehler beim Senden eines Spark-Jobs mit Docker-Laufzeit zu vermeiden.
- In Amazon EMR 6.0.0 ist der Speichermodus HBase auf Amazon S3 vom Problem [HBASE-24286](#) betroffen. HBaseMaster kann nicht initialisiert werden, wenn der Cluster mit vorhandenen S3-Daten erstellt wird.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop` Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.30.1

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.30.1. Änderungen beziehen sich auf Version 5.30.0.

Erstveröffentlichungsdatum: 30. Juni 2020

Letzte Aktualisierung: 24. August 2020

Änderungen, Verbesserungen und behobene Probleme

- Neuere EMR Amazon-Versionen beheben das Problem mit einem niedrigeren Limit für „Max. Anzahl geöffneter Dateien“ für ältere Versionen AL2 in AmazonEMR. Die EMR Amazon-Versionen 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 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Es wurde ein Problem behoben, bei dem der Instance-Controller-Prozess unendlich viele Prozesse ausgelöst hat.
- Es wurde ein Problem behoben, bei dem Hue keine Hive-Abfrage ausführen konnte, die Meldung „Datenbank ist gesperrt“ angezeigt wurde und die Ausführung von Abfragen verhindert wurde.
- Es wurde ein Spark-Problem behoben, durch das mehr Aufgaben gleichzeitig auf dem Cluster ausgeführt werden konnten. EMR
- Es wurde ein Problem mit dem Jupyter Notebook behoben, das den Fehler „Zu viele Dateien öffnen“ auf dem Jupyter-Server verursachte.
- Ein Problem mit den Startzeiten von Clustern wurde behoben.

Neue Features

- Die persistenten Anwendungsschnittstellen der Tez-Benutzeroberfläche und des YARN Timeline-Servers sind in den EMR Amazon-Versionen 6.x und 5.30.1 und EMR höher verfügbar. Durch den

Linkzugriff mit einem Klick auf den persistenten Anwendungsverlauf können Sie schnell auf den Jobverlauf zugreifen, ohne einen Web-Proxy über eine Verbindung einrichten zu müssen. SSH Protokolle für aktive und beendete Cluster sind 30 Tage nach Ende der Anwendung verfügbar. Weitere Informationen finden Sie unter [Benutzerschnittstellen für persistente Anwendungen anzeigen](#) im Amazon EMR Management Guide.

- EMR Notebook Execution APIs ist verfügbar, um EMR Notebooks über ein Skript oder eine Befehlszeile auszuführen. Durch die Möglichkeit, EMR Notebook-Ausführungen ohne die AWS Konsole zu starten, zu beenden, aufzulisten und zu beschreiben, können Sie ein Notebook programmgesteuert steuern. EMR Mithilfe einer parametrisierten Notebook-Zelle können Sie verschiedene Parameterwerte an ein Notebook übergeben, ohne für jeden neuen Satz von Parameterwerten eine Kopie des Notebooks erstellen zu müssen. [Siehe Aktionen. EMR API](#) Beispielcode finden Sie unter [Beispielbefehle zur programmgesteuerten Ausführung von EMR Notebooks](#).

Bekannte Probleme

- Niedrigerer Grenzwert für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMRAmazon-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von ofAmazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-StandardEinstellung eine Ulimit-StandardEinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latestAmazon AMI Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- EMRNotizbücher

Die Funktion, mit der Sie Kernel und zusätzliche Python-Bibliotheken auf dem primären Clusterknoten installieren können, ist in EMR Version 5.30.1 standardmäßig deaktiviert. Weitere Informationen zu diesem Feature finden Sie unter [Installation von Kernels und Python-Bibliotheken auf einem Cluster-Primärknoten](#).

Um das Feature zu aktivieren, gehen Sie wie folgt vor:

1. Stellen Sie sicher, dass die der Servicerolle für EMR Notebooks zugeordnete Berechtigungsrichtlinie die folgende Aktion zulässt:

```
elasticmapreduce:ListSteps
```

Weitere Informationen finden Sie unter [Servicerolle für EMR Notebooks](#).

2. Verwenden Sie den AWS CLI , um einen Schritt auf dem Cluster auszuführen, der EMR Notebooks einrichtet, wie im folgenden Beispiel gezeigt. Ersetzen *us-east-1* mit der Region, in der sich Ihr Cluster befindet. Weitere Informationen finden Sie unter [Hinzufügen von Schritten zu einem Cluster mithilfe 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"]
```

- **Verwaltete Skalierung**

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den EMR Amazon-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Job Presto nicht benötigt.

- **Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung**

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde.

Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 5.30.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.30.0. Änderungen beziehen sich auf Version 5.29.0.

Erste Version: 13. Mai 2020

Letzte Aktualisierung: 25. Juni 2020

Upgrades

- Auf Version AWS SDK for Java 1.11.759 aktualisiert
- Amazon SageMaker Spark wurde SDK auf Version 1.3.0 aktualisiert
- EMRRecord Server wurde auf Version 1.6.0 aktualisiert
- Flink auf Version 1.10.0 aktualisiert
- Ganglia auf Version 3.7.2 aktualisiert
- Auf Version HBase 1.4.13 aktualisiert
- Hudi auf Version 0.5.2-incubating aktualisiert
- Hue auf Version 4.6.0 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Livy auf Version 0.7.0-incubating aktualisiert
- Oozie auf Version 5.2.0 aktualisiert
- Presto auf Version 0.232 aktualisiert
- Spark auf Version 2.4.5 aktualisiert
- Aktualisierte Konnektoren und Treiber: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; DynamoDB Connector 4.14.0 EMR

Neue Features

- EMRNotebooks — Bei Verwendung mit EMR Clustern, die mit 5.30.0 erstellt wurden, werden Notebook-Kernel auf dem Cluster ausgeführt. EMR Dies verbessert die Notebook-Leistung und

ermöglicht es Ihnen, Kernel zu installieren und anzupassen. Sie können Python-Bibliotheken auch auf dem Cluster-Primärknoten installieren. Weitere Informationen finden Sie unter [Installation und Verwendung von Kernels and Libraries im Management Guide](#). EMR

- **Verwaltete Skalierung** — Mit EMR Amazon-Version 5.30.0 und höher können Sie EMR Managed Scaling aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster je nach Arbeitslast automatisch zu erhöhen oder zu verringern. Amazon wertet EMR kontinuierlich Cluster-Metriken aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster im Hinblick auf Kosten und Geschwindigkeit optimieren. Weitere Informationen finden Sie unter [Scaling Cluster Resources](#) im Amazon EMR Management Guide.
- **In Amazon S3 gespeicherte Protokolldateien verschlüsseln** — Mit Amazon EMR Version 5.30.0 und höher können Sie in Amazon S3 gespeicherte Protokolldateien mit einem AWS KMS vom Kunden verwalteten Schlüssel verschlüsseln. Weitere Informationen finden Sie unter [Verschlüsseln von in Amazon S3 gespeicherten Protokolldateien](#) im Amazon EMR Management Guide.
- **Amazon Linux 2-Unterstützung** — In EMR Version 5.30.0 und höher, EMR uses Amazon Linux 2-Betriebssystem. Das neue benutzerdefinierte AMIs (Amazon Machine Image) muss auf the Amazon Linux 2 basieren AMI. Weitere Informationen finden Sie unter [Benutzerdefiniert verwenden AMI](#).
- **Presto Graceful Auto Scale** — Für EMR Cluster, die 5.30.0 verwenden, kann ein Timeout für die automatische Skalierung festgelegt werden, sodass Presto-Aufgaben Zeit haben, ihre Ausführung zu beenden, bevor ihr Knoten außer Betrieb genommen wird. Weitere Informationen finden Sie unter [Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden](#).
- **Erstellung einer Flotteninstanz mit neuer Zuweisungsstrategieoption** — Eine neue Zuweisungsstrategieoption ist in Version 5.12.1 und höher verfügbar. EMR Sie bietet eine schnellere Cluster-Bereitstellung, eine genauere Spot-Zuweisung und weniger Unterbrechungen von Spot Instances. Aktualisierungen für nicht standardmäßige EMR Servicerollen sind erforderlich. Sehen Sie unter [Konfigurieren von Instance-Flotten](#).
- **Befehle sudo systemctl stop und sudo systemctl start** — In EMR Version 5.30.0 und höher, die das Betriebssystem use Amazon Linux 2 verwendet, und Befehle, um Dienste neu zu starten. EMR sudo systemctl stop sudo systemctl start Weitere Informationen finden Sie unter [Wie starte ich einen Service bei Amazon neu EMR?](#) .

Änderungen, Verbesserungen und behobene Probleme

- EMR Version 5.30.0 installiert Ganglia nicht standardmäßig. Sie können Ganglia explizit für die Installation auswählen, wenn Sie einen Cluster erstellen.
- Spark-Leistungsoptimierungen.

- Presto-Leistungsoptimierungen.
- Python 3 ist der Standard für EMR Amazon-Version 5.30.0 und höher.
- Die standardmäßige verwaltete Sicherheitsgruppe für den Zugriff auf Services in privaten Subnetzen wurde mit neuen Regeln aktualisiert. Wenn Sie benutzerdefinierte Sicherheitsgruppe für den Servicezugriff verwenden, müssen Sie dieselben Regeln wie die standardmäßige verwaltete Sicherheitsgruppe einschließen. Weitere Informationen finden Sie unter [Amazon EMR - Managed Security Group for Service Access \(Private Subnets\)](#). Wenn Sie eine benutzerdefinierte Servicerolle für Amazon verwenden EMR, müssen Sie die entsprechenden Berechtigungen erteilen, `ec2:describeSecurityGroups` damit überprüft werden EMR kann, ob die Sicherheitsgruppen korrekt erstellt wurden. Wenn Sie `EMR_DefaultRole` verwenden, ist diese Berechtigung bereits in der standardmäßigen verwalteten Richtlinie enthalten.

Bekannte Probleme

- Niedrigeres Limit für die maximale Anzahl geöffneter Dateien bei älteren Versionen AL2 [in neueren Versionen behoben]. EMR Amazon-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Geöffnete Dateien“ haben, wenn Amazon-Cluster mit der Standardeinstellung erstellt werden. EMR AMI Die EMR Amazon-Versionen 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 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen AMI hat die EMR Amazon-Standardeinstellung eine `ulimit`-Standardeinstellung von 4096 für „Max. Geöffnete Dateien“, was unter dem Dateilimit von 65536 in Linux 2 liegt. latest Amazon AMI Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, EMR verfügt Amazon über ein Bootstrap Action (BA) -Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere EMR Amazon-Version verwenden, die nicht über die permanente Lösung für dieses Problem verfügt, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/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
```

- **Verwaltete Skalierung**

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den EMR Amazon-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Job Presto nicht benötigt.

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Die Standard-Datenbank-Engine für Hue 4.6.0 ist SQLite, was zu Problemen führt, wenn Sie versuchen, Hue mit einer externen Datenbank zu verwenden. Um dieses Problem zu beheben, setzen Sie `engine` in Ihrer `hue.ini` Konfigurationsklassifizierung auf `mysql`. Dieses Problem wurde in EMR Amazon-Version 5.30.1 behoben.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsspeicherorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den EMR Amazon-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
 - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
 - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
 - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen Wert von UTF -8, der kleiner als das Zeichen (U+002F) ist. / Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [Kodierungstabelle UTF -8](#) und Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

Version 5.29.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.29.0. Änderungen beziehen sich auf Version 5.28.1.

Erste Version: 17. Januar 2020

Upgrades

- Auf Version AWS SDK for Java 1.11.682 aktualisiert
- Upgrade von Hive auf Version 2.3.6
- Flink wurde auf Version 1.9.1 aktualisiert
- EmrFS wurde auf Version 2.38.0 aktualisiert
- EMRDynamoDB Connector wurde auf Version 4.13.0 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Spark-Leistungsoptimierungen.
- EMRFS
 - Management Guide aktualisiert die Standardeinstellungen von emrfs-site.xml für eine konsistente Ansicht.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.28.1

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.28.1. Änderungen beziehen sich auf Version 5.28.0.

Erste Version: 10. Januar 2020

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Spark-Kompatibilitätsprobleme behoben.
- CloudWatch Metriken
 - Fehler bei der Veröffentlichung von Amazon CloudWatch Metrics auf einem EMR Cluster mit mehreren Primärknoten behoben.
- Deaktivierte Protokollnachricht
 - Falsche Protokollnachricht „... verwendet die alte Version (<4.5.8) des Apache-HTTP-Clients“ deaktiviert.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.28.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.28.0. Änderungen beziehen sich auf Version 5.27.0.

Erste Version: 12. November 2019

Upgrades

- Upgrade von Flink auf Version 1.9.0
- Upgrade von Hive auf Version 2.3.6
- Auf Version MXNet 1.5.1 aktualisiert
- Upgrade von Phoenix auf Version 4.14.3
- Upgrade von Presto auf Version 0.227
- Upgrade von Zeppelin auf Version 0.8.2

Neue Features

- [Apache Hudi](#) kann jetzt von Amazon installiert EMR werden, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Hudi](#).

- (25. November 2019) Sie können jetzt auf Wunsch mehrere Schritte gleichzeitig ausführen, um die Cluster-Nutzung zu verbessern und Kosten zu sparen. Außerdem können Sie sowohl anstehende als auch laufende Schritte stornieren. Weitere Informationen finden Sie unter [Arbeiten mit Schritten mithilfe der AWS CLI AND-Konsole](#).
- (3. Dezember 2019) Sie können jetzt EMR Cluster auf erstellen und ausführen AWS Outposts. AWS Outposts ermöglicht native AWS Dienste, Infrastrukturen und Betriebsmodelle in lokalen Einrichtungen. In AWS Outposts Umgebungen können Sie dieselben AWS APIs Tools und dieselbe Infrastruktur verwenden wie in der AWS Cloud. Weitere Informationen finden Sie unter [EMRCluster on AWS Outposts](#).
- (11. März 2020) Ab EMR Amazon-Version 5.28.0 können Sie EMR Amazon-Cluster in einem Local Zones-Subnetz als logische Erweiterung einer AWS Region, die AWS Local Zones unterstützt, erstellen und ausführen. Eine lokale Zone ermöglicht es, EMR Amazon-Funktionen und eine Untergruppe von AWS Diensten, wie Rechen- und Speicherdienste, näher an den Benutzern zu platzieren, wodurch der Zugriff auf lokal ausgeführte Anwendungen mit sehr geringer Latenz ermöglicht wird. Eine Liste der verfügbaren Local Zones finden Sie unter [AWS Local Zones](#). Informationen zum Zugriff auf verfügbare AWS Local Zones finden Sie unter [Regionen, Availability Zones und Local Zones](#).

Local Zones unterstützen derzeit keine EMR Amazon-Notebooks und unterstützen keine direkten Verbindungen zu Amazon EMR über den VPC Schnittstellenendpunkt (AWS PrivateLink).

Änderungen, Verbesserungen und behobene Probleme

- Erweiterter Anwendungssupport für Cluster mit hoher Verfügbarkeit
 - Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR Cluster mit mehreren Primärknoten](#) im Amazon EMR Management Guide.
- Spark
 - Leistungsoptimierungen
- Hive
 - Leistungsoptimierungen
- Presto
 - Leistungsoptimierungen

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Veröffentlichung 5.27.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.27.0. Änderungen beziehen sich auf Version 5.26.0.

Erste Version: 23. September 2019

Upgrades

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1,14.0
- Konnektoren und Treiber:
 - DynamoDB Connector 4.12.0

Neue Features

- (24. Oktober 2019) Die folgenden neuen Funktionen in EMR Notizbüchern sind in allen EMR Amazon-Versionen verfügbar.
 - Sie können jetzt Git-Repositorys mit EMR Notizbüchern verknüpfen, um Ihre Notizbücher in einer versionskontrollierten Umgebung zu speichern. Sie können Code mit Kollegen teilen und vorhandene Jupyter Notebooks über Remote-Git-Repositorys wiederverwenden. Weitere Informationen finden Sie unter [Git-Repositorys mit Amazon EMR Notebooks verknüpfen](#) im Amazon EMR Management Guide.
 - Das [Hilfsprogramm nbdime](#) ist jetzt in EMR Notebooks verfügbar, um das Vergleichen und Zusammenführen von Notizbüchern zu vereinfachen.
 - EMRNotebooks werden jetzt unterstützt. JupyterLab JupyterLab ist eine webbasierte interaktive Entwicklungsumgebung, die vollständig mit Jupyter-Notebooks kompatibel ist. Sie können jetzt wählen, ob Sie Ihr Notizbuch entweder im JupyterLab Jupyter-Notebook-Editor öffnen möchten.
- (30. Oktober 2019) Mit den EMR Amazon-Versionen 5.25.0 und höher können Sie über die Cluster-Übersichtsseite oder den Tab Anwendungsverlauf in der Konsole eine Verbindung zur Benutzeroberfläche des Spark-Verlaufs herstellen. Anstatt einen Web-Proxy über eine SSH Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbetzeroberflächen](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- EMRAmazon-Cluster mit mehreren Primärknoten

- Sie können Flink auf einem EMR Amazon-Cluster mit mehreren Primärknoten installieren und ausführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen und Features](#).
- Sie können HDFS transparente Verschlüsselung auf einem EMR Amazon-Cluster mit mehreren Primärknoten konfigurieren. Weitere Informationen finden Sie unter [HDFSTransparente Verschlüsselung auf EMR Clustern mit mehreren Primärknoten](#).
- Sie können jetzt die Konfiguration von Anwendungen ändern, die auf einem EMR Amazon-Cluster mit mehreren Primärknoten ausgeführt werden. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).
- EMRAmazon-DynamoDB-Konnektor
 - Amazon EMR -DynamoDB Connector unterstützt jetzt die folgenden DynamoDB-Datentypen: boolean, list, map, item, null. Weitere Informationen finden Sie unter [Eine Hive-Tabelle zur Ausführung von Hive-Befehlen einrichten](#).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.26.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.26.0. Änderungen beziehen sich auf Version 5.25.0.

Erste Version: 8. August 2019

Letzte Aktualisierung: 19. August 2019

Upgrades

- AWS SDK for Java 1.11.595
- HBase1.4,10
- Phoenix 4.14.2
- Konnektoren und Treiber:
 - DynamoDB Connector 4.11.0
 - MariaDB Connector 2.4.2
 - Amazon Redshift JDBC Redshift-Treiber 1.2.32.1056

Neue Features

- (Beta) Mit Amazon EMR 5.26.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration bietet einen detaillierten Zugriff auf Spaltenebene auf Datenbanken und Tabellen im Glue-Datenkatalog. AWS ermöglicht auch föderiertes Single Sign-On für EMR Notebooks oder Apache Zeppelin von einem Unternehmensidentitätssystem aus. Weitere Informationen finden Sie unter [Integration von Amazon EMR mit AWS Lake Formation \(Beta\)](#).
- (19. August 2019) Amazon EMR Block Public Access ist jetzt in allen EMR Amazon-Versionen verfügbar, die Sicherheitsgruppen unterstützen. Den öffentlichen Zugriff blockieren ist eine

kontoweite Einstellung, die für jede AWS Region gilt. Den öffentlichen Zugriff blockieren verhindert, dass ein Cluster gestartet wird, wenn eine dem Cluster zugeordnete Sicherheitsgruppe über eine Regel verfügt, die eingehenden Datenverkehr von IPv4 0.0.0.0/0 oder IPv6:: /0 (öffentlicher Zugriff) auf einem Port zulässt, sofern kein Port als Ausnahme angegeben ist. Port 22 ist standardmäßig eine Ausnahme. Weitere Informationen finden Sie unter [Using Amazon EMR Block Public Access](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- EMRNotizbücher
 - Ab Version EMR 5.26.0 unterstützt EMR Notebooks zusätzlich zu den Standard-Python-Bibliotheken auch Python-Bibliotheken für Notebooks. Sie können Bibliotheken für Notebooks vom Notebook-Editor aus installieren, ohne einen Cluster neu erstellen oder ein Notebook erneut an einen Cluster anhängen zu müssen. Notebook-bezogene Bibliotheken werden in einer virtuellen Python-Umgebung erstellt und gelten daher nur für die aktuelle Notebook-Sitzung. Auf diese Weise können Sie Abhängigkeiten zwischen Notebooks isolieren. Weitere Informationen finden Sie unter [Using Notebook Scoped Libraries](#) im Amazon EMR Management Guide.
- EMRFS
 - Sie können eine ETag Bestätigungsfunktion (Beta) aktivieren, indem Sie auf `einstellenfs.s3.consistent.metadata.etag.verification.enabled.true` Mit dieser Funktion überprüft Amazon EMRFS S3, ob ETags es sich bei den gelesenen Objekten um die neueste verfügbare Version handelt. Diese Funktion ist hilfreich für read-after-update Anwendungsfälle, in denen Dateien auf Amazon S3 überschrieben werden, während derselbe Name beibehalten wird. Diese ETag Überprüfungsfunktion funktioniert derzeit nicht mit S3 Select. Weitere Informationen finden Sie unter [Konfigurieren der konsistenten Ansicht](#).
- Spark
 - Die folgenden Optimierungen sind jetzt standardmäßig aktiviert: dynamisches Bereinigen von Partitionen DISTINCT zuvor INTERSECT, Verbesserungen bei der Ableitung von SQL Planstatistiken für JOIN nachfolgende DISTINCT Abfragen, Reduzierung skalarer Unterabfragen, optimierte Neureihenfolge von Verknüpfungen und Bloom-Filter-Join. Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).
 - Die Codegenerierung für die gesamte Phase für Sort Merge Join wurde verbessert.
 - Die Wiederverwendung von Abfragefragmenten und Unterabfragen wurde verbessert.
 - Verbesserungen bei der Vorabzuweisung von Executoren beim Start von Spark.

- Bloom-Filter-Joins werden nicht mehr angewendet, wenn die kleinere Seite der Verknüpfung einen Broadcast-Hinweis enthält.
- Tez
 - Ein Problem mit Tez wurde behoben. Tez UI funktioniert jetzt auf einem EMR Amazon-Cluster mit mehreren Primärknoten.

Bekannte Probleme

- Die verbesserten Funktionen zur Codegenerierung in der gesamten Phase für Sort Merge Join können den Speicherdruck erhöhen, wenn sie aktiviert sind. Diese Optimierung verbessert die Leistung, kann jedoch zu Wiederholungsversuchen oder Fehlschlägen führen, wenn der nicht so eingestellt `spark.yarn.executor.memoryOverheadFactor` ist, dass er genügend Speicher zur Verfügung stellt. Um dieses Feature zu deaktivieren, setzen Sie `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` auf „false“.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.25.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.25.0. Änderungen beziehen sich auf Version 5.24.1.

Erste Version: 17. Juli 2019

Letzte Aktualisierung: 30. Oktober 2019

Amazon EMR 5,25,0

Upgrades

- 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

Neue Features

- (30. Oktober 2019) Ab EMR Amazon-Version 5.25.0 können Sie über die Cluster-Übersichtsseite oder den Tab Anwendungsverlauf in der Konsole eine Verbindung zur Benutzeroberfläche des Spark-Verlaufs herstellen. Anstatt einen Web-Proxy über eine SSH Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete

Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbenuzoberflächen](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Die Leistung einiger Verknüpfungen wurde verbessert, indem Bloom-Filter zum Vorfiltern von Eingaben verwendet wurden. Die Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.bloomFilterJoin.enabled` auf `true` gesetzt wird.
 - Die Leistung bei der Gruppierung nach Spalten vom Typ Zeichenfolge wurde verbessert.
 - Die standardmäßige Speicher- und Kernkonfiguration von Spark Executor für R4-Instance-Typen für Cluster ohne Installation wurde verbessert. HBase
 - Es wurde ein früheres Problem mit dem Feature zum dynamischen Bereinigen von Partitionen behoben, bei dem sich die bereinigte Tabelle auf der linken Seite des Joins befinden musste.
 - DISTINCT vor der INTERSECT Optimierung verbessert und gilt nun auch für weitere Fälle, in denen Aliase involviert sind.
 - Die Inferenz von SQL Planstatistiken für Abfragen, JOIN denen gefolgt von DISTINCT Abfragen, wurde verbessert. Diese Verbesserung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.statsImprovements.enabled` auf `true` gesetzt wird. Diese Optimierung ist für das Feature „Unterscheiden vor Überschneiden“ erforderlich und wird automatisch aktiviert, wenn `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true` gesetzt ist.
 - Optimierte Verbindungsreihenfolge basierend auf Tabellengröße und Filtern. Diese Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` auf `true` gesetzt wird.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- EMRFS
 - Die EMRFS Einstellung, `fs.s3.buckets.create.enabled`, ist jetzt standardmäßig deaktiviert. Bei Tests haben wir festgestellt, dass die Deaktivierung dieser Einstellung die Leistung verbessert und die unbeabsichtigte Erstellung von S3-Buckets verhindert. Wenn Ihre Anwendung auf dieser Funktionalität basiert, können Sie sie aktivieren, indem Sie die Eigenschaft `true` in der Konfigurationsklassifizierung von `fs.s3.buckets.create.enabled`

auf `emrfs-site` setzen. Weitere Informationen finden Sie unter [Angeben einer Konfiguration beim Erstellen eines Clusters](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
 - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
 - Es wurde eine Option hinzugefügt, um die EBS Verschlüsselung mit Version 5.24.0 und höher zu aktivieren. Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten Datenträgers AMI zur Verschlüsselung des Root-Geräte-Volumens.
 - Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#) im Amazon EMR Management Guide.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop` Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.24.1

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.24.1. Änderungen beziehen sich auf Version 5.24.0.

Erste Version: 26. Juni 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.24.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.24.0. Änderungen beziehen sich auf Version 5.23.0.

Erste Version: 11. Juni 2019

Letzte Aktualisierung: 5. August 2019

Upgrades

- 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
- Konnektoren und Treiber:

- DynamoDB Connector 4.9.0
- MariaDB Connector 2.4.1
- Amazon Redshift JDBC Redshift-Treiber 1.2.27.1051

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Es wurde eine Optimierung für das dynamische Bereinigen von Partitionen hinzugefügt. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.dynamicPartitionPruning.enabled` auf `true`.
 - Die Leistung von INTERSECT-Anfragen wurde verbessert. Diese Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true`.
 - Es wurde eine Optimierung hinzugefügt, um skalare Unterabfragen mit Aggregaten, die dieselbe Beziehung verwenden, zu vereinfachen. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` auf `true`.
 - Die Codegenerierung für die gesamte Phase wurde verbessert.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
 - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
 - Es wurde eine Option hinzugefügt, um die Verschlüsselung zu aktivieren. EBS Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten `VolumesAMI`, um das Root-Geräte-Volumen zu verschlüsseln.
 - Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#) im Amazon EMR Management Guide.

Bekanntes Problem

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.23.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.23.0. Änderungen beziehen sich auf Version 5.22.0.

Erste Version: 01. April 2019

Letzte Aktualisierung: 30. April 2019

Upgrades

- AWS SDK for Java 1.11.519

Neue Features

- (30. April 2019) Mit Amazon EMR 5.23.0 und höher können Sie einen Cluster mit drei primären Knoten starten, um die Hochverfügbarkeit von Anwendungen wie YARN Resource Manager, Spark HDFS NameNode, Hive und Ganglia zu unterstützen. Der Primärknoten ist mit diesem Feature keine potenzielle einzelne Fehlerquelle mehr. Wenn einer der Primärknoten ausfällt, wechselt Amazon EMR automatisch zu einem Standby-Primärknoten und ersetzt den ausgefallenen Primärknoten durch einen neuen mit derselben Konfiguration und denselben Bootstrap-Aktionen. Weitere Informationen finden Sie unter [Primärknoten planen und konfigurieren](#).

Bekannte Probleme

- Tez UI (In EMR Amazon-Version 5.26.0 behoben)

Die Tez-Benutzeroberfläche funktioniert nicht auf einem EMR Cluster mit mehreren Primärknoten.

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
 - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

- Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.22.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.22.0. Änderungen beziehen sich auf Version 5.21.0.

Important

Ab EMR Amazon-Version 5.22.0 EMR verwendet Amazon AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Frühere EMR Amazon-Versionen verwenden in einigen Fällen AWS Signature Version 2, sofern in den Versionshinweisen nicht angegeben ist, dass ausschließlich Signature Version 4 verwendet wird. Weitere Informationen finden Sie unter [Authentifizieren von Anfragen \(AWS Signature Version 4\)](#) und [Authentifizieren von Anfragen \(AWS Signature Version 2\)](#) im Amazon Simple Storage Service Developer Guide.

Erste Version: 20. März 2019

Upgrades

- Flink 1.7.1
- HBase 1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Konnektoren und Treiber:
 - DynamoDB Connector 4.8.0
 - MariaDB Connector 2.2.6
 - Amazon Redshift JDBC Redshift-Treiber 1.2.20.1043

Neue Features

- Die EBS Standardkonfiguration für EC2 Instance-Typen mit -only-Speicher wurde geändert. EBS Wenn Sie einen Cluster mit Amazon EMR Version 5.22.0 und höher erstellen, erhöht sich die EBS Standardspeichermenge je nach Größe der Instance. Darüber hinaus haben wir den erhöhten Speicherplatz auf mehrere Volumes aufgeteilt, um die Leistung zu erhöhen IOPS. Wenn Sie eine

andere EBS Instance-Speicherkonfiguration verwenden möchten, können Sie diese angeben, wenn Sie einen EMR Cluster erstellen oder Knoten zu einem vorhandenen Cluster hinzufügen. Weitere Informationen zur Speichermenge und Anzahl der standardmäßig zugewiesenen Volumes für jeden Instance-Typ finden Sie unter [EBSStandardspeicher für Instances](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Es wurde eine neue Konfigurationseigenschaft für Spark aufYARN, eingeführt `spark.yarn.executor.memoryOverheadFactor`. Der Wert dieser Eigenschaft ist ein Skalierungsfaktor, der den Wert des Speicher-Overheads auf einen Prozentsatz des Executor-Speichers festlegt, mit einem Minimum von 384 MB. Wenn der Speicher-Overhead explizit mit `spark.yarn.executor.memoryOverhead` festgelegt wird, hat diese Eigenschaft keine Auswirkung. Der Standardwert ist `0.1875`, was 18.75 % entspricht. Dieser Standard für Amazon EMR lässt mehr Platz in YARN Containern für den Executor-Speicheraufwand übrig als der intern von Spark festgelegte Standard von 10%. Der EMR Amazon-Standard von 18,75% ergab empirisch weniger speicherbedingte Ausfälle in -DS-Benchmarks. TPC
 - [-26316 wurde zurückportiert, um die Leistung zu verbessern. SPARK](#)
- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS

Bekannte Probleme

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
 - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.21.1

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.21.1. Änderungen beziehen sich auf Version 5.21.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.21.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.21.0. Änderungen beziehen sich auf Version 5.20.0.

Veröffentlichungsdatum: 18. Februar 2019

Letzte Aktualisierung: 3. April 2019

Upgrades

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

Neue Features

- (3. April 2019) Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Sie tun dies, indem Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Änderungen, Verbesserungen und behobene Probleme

- Zeppelin
 - Rückportiert [ZEPPELIN-3878](#).

Bekannte Probleme

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
 - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
 - Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie unter `http://` eine Verbindung zur Tez-Benutzeroberfläche herstellen `MasterDNS:8080/tez-ui` über eine SSH Verbindung zum primären Clusterknoten wird der Fehler „Adaptervorgang fehlgeschlagen - Timeline-Server () ATS ist außer Reichweite“ angezeigt. Entweder ist er

ausgefallen oder nicht aktiviert“ CORS wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server sendet und localhost nicht den Hostnamen des primären Knotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.20.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.20.0. Änderungen beziehen sich auf Version 5.19.0.

Veröffentlichungsdatum: 18. Dezember 2018

Letzte Aktualisierung: 22. Januar 2019

Upgrades

- 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

Neue Features

- (22. Januar 2019) Kerberos in Amazon EMR wurde verbessert und unterstützt nun die Authentifizierung von Principals von einem externen System aus. KDC Dadurch wird die Prinzipalverwaltung zentralisiert, da sich mehrere Cluster einen einzigen externen Cluster teilen können. KDC Darüber hinaus KDC kann das externe System eine bereichsübergreifende

Vertrauensstellung mit einer Active Directory-Domäne haben. Auf diese Weise können alle Cluster Prinzipale von Active Directory authentifizieren. Weitere Informationen finden Sie unter [Verwenden der Kerberos-Authentifizierung](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Standard Amazon Linux AMI für Amazon EMR
 - Das Python3-Paket wurde von Python 3.4 auf 3.6 aktualisiert.
- Der EMRFS S3-optimierte Committer
 - Der EMRFS S3-optimierte Committer ist jetzt standardmäßig aktiviert, was die Schreibleistung verbessert. Weitere Informationen finden Sie unter [Verwenden Sie den EMRFS S3-optimierten Committer](#).
- Hive
 - [Rückportiert -16686. HIVE](#)
- Mit Spark und Hive Glue
 - In Version EMR 5.20.0 oder höher wird die parallel Partitionsbereinigung automatisch für Spark und Hive aktiviert, wenn AWS Glue Data Catalog als Metastore verwendet wird. Diese Änderung reduziert die Zeit für die Abfrageplanung erheblich, da mehrere Anfragen parallel ausgeführt werden, um Partitionen abzurufen. Die Gesamtzahl der Segmente, die gleichzeitig ausgeführt werden können, liegt zwischen 1 und 10. Der Standardwert ist 5, was eine empfohlene Einstellung ist. Sie können dies ändern, indem Sie die Eigenschaft `aws.glue.partition.num.segments` in der `hive-site`-Konfigurationsklassifizierung angeben. Wenn eine Drosselung auftritt, können Sie das Feature ausschalten, indem Sie den Wert auf 1 ändern. Weitere Informationen finden Sie unter [AWS -Glue-Segmentstruktur](#).

Bekannte Probleme

- Farbton (In EMR Amazon-Version 5.24.0 behoben)
 - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab EMR Amazon-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
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',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie mit der Befehlszeile des primären Knotens eine ConnectSSH.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
- Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie unter `http://` eine Verbindung zur Tez-Benutzeroberfläche herstellen `MasterDNS:8080/tez-ui` über eine SSH Verbindung zum primären Clusterknoten wird der Fehler „Adaptervorgang fehlgeschlagen - Timeline-Server () ATS ist außer Reichweite“ angezeigt. Entweder ist er ausgefallen oder nicht aktiviert“ CORS wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server sendet und `localhost` nicht den Hostnamen des primären Knotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden.

HDFS

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in EMR Amazon-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie Scale-Down oder Step-Submission auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop Benutzer des führenden Primärknotens des EMR Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

Version 5.19.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.19.0. Änderungen beziehen sich auf Version 5.18.0.

Erste Version: 7. November 2018

Letzte Aktualisierung: 19. November 2018

Upgrades

- 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

Neue Features

- (19. November 2018) EMR Notebooks ist eine verwaltete Umgebung, die auf Jupyter Notebook basiert. Es unterstützt Spark-Magic-Kernel für Spark PySparkSQL, Spark R und Scala. EMRNotebooks können mit Clustern verwendet werden, die mit Amazon EMR Version 5.18.0 und höher erstellt wurden. Weitere Informationen finden Sie unter [Using EMR Notebooks](#) im Amazon EMR Management Guide.
- Der EMRFS S3-optimierte Committer ist verfügbar, wenn Sie Parquet-Dateien mit Spark und schreiben. EMRFS Dieser Committer verbessert die Schreibleistung. Weitere Informationen finden Sie unter [Verwenden Sie den EMRFS S3-optimierten Committer](#).

Änderungen, Verbesserungen und behobene Probleme

- YARN
 - Die Logik, die den Anwendungsmasterprozess auf die Ausführung auf Core-Knoten beschränkt, wurde geändert. Diese Funktion verwendet jetzt die Funktion und Eigenschaften von YARN Knotenbeschriftungen in den Klassifizierungen `yarn-site` und `capacity-scheduler` der Konfiguration. Weitere Informationen finden Sie unter <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- Standard Amazon Linux AMI für Amazon EMR
 - `ruby18`, `php56` und `gcc48` sind nicht mehr standardmäßig installiert. Diese können auf Wunsch mit `yum` installiert werden.

- Das `aws-sdk` Ruby Gem ist standardmäßig nicht mehr installiert. Es kann, falls gewünscht, mit `gem install aws-sdk` installiert werden. Bestimmte Komponenten können ebenfalls installiert werden. Beispiel, `gem install aws-sdk-s3`.

Bekannte Probleme

- **EMRNotebooks** — Unter bestimmten Umständen kann es vorkommen, dass der Notebook-Editor bei mehreren geöffneten Notebook-Editoren keine Verbindung zum Cluster herstellen kann. Löschen Sie in diesem Fall die Browser-Cookies und öffnen Sie die Notebook-Editoren erneut.
- **CloudWatch ContainerPending Metrische und automatische Skalierung** — (In 5.20.0 behoben) Amazon gibt EMR möglicherweise einen negativen Wert für `ContainerPending` Wenn `ContainerPending` in einer Auto-Scaling-Regel verwendet wird, verhält sich das Auto Scaling nicht wie erwartet. Vermeiden Sie die Verwendung von `ContainerPending` mit automatischer Skalierung.
- In den EMR Amazon-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN Knotenbezeichnungen in einem Verzeichnis gespeichert. HDFS In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARNKnotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von vermieden werden. HDFS

Version 5.18.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.18.0. Änderungen beziehen sich auf Version 5.17.0.

Veröffentlichungsdatum: 24. Oktober 2018

Upgrades

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Neue Features

- Ab Amazon EMR 5.18.0 können Sie das Amazon EMR Artifact Repository verwenden, um Ihren Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die in bestimmten Amazon-Versionen verfügbar sind. EMR Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Änderungen, Verbesserungen und behobene Probleme

- Hive
 - Unterstützung für S3 Select hinzugefügt. Weitere Informationen finden Sie unter [Verwendung von S3 Select mit Hive zur Verbesserung der Leistung](#).
- Presto
 - Unterstützung für [S3 Select](#) Pushdown hinzugefügt. Weitere Informationen finden Sie unter [S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden](#).
- Spark
 - Die standardmäßige log4j-Konfiguration für Spark wurde dahingehend geändert, dass Container-Logs für Spark-Streaming-Aufträge stündlich angezeigt werden. Dadurch wird verhindert, dass Logs für Spark-Streaming-Aufträge mit langer Laufzeit gelöscht werden.

Version 5.17.1

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.17.1. Änderungen beziehen sich auf Version 5.17.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

Version 5.17.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.17.0. Änderungen beziehen sich auf Version 5.16.0.

Erste Version: 30. August 2018

Upgrades

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

Neue Features

- Unterstützung für Tensorflow hinzugefügt. Weitere Informationen finden Sie unter [TensorFlow](#).

Änderungen, Verbesserungen und behobene Probleme

- JupyterHub
 - Unterstützung für Notebook-Persistenz in Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter [Konfigurieren von Persistenz für Notebooks in Amazon S3](#).
- Spark
 - Unterstützung für [S3 Select](#) hinzugefügt. Weitere Informationen finden Sie unter [S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden](#).
- Die Probleme mit den Cloudwatch-Metriken und der automatischen Skalierungsfunktion in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 wurden behoben.

Bekannte Probleme

- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Wenn Sie ein benutzerdefiniertes Amazon Linux verwenden, das auf einem Amazon Linux AMI mit einem Erstellungsdatum vom 11.08.2018 AMI basiert, kann der Oozie-Server nicht gestartet werden. Wenn Sie Oozie verwenden, erstellen Sie eine benutzerdefinierte Version, die auf einer Amazon AMI Linux-ID mit einem anderen Erstellungsdatum AMI basiert. Sie können den folgenden AWS CLI Befehl verwenden, um eine Liste mit Images IDs für alle HVM Amazon Linux-Versionen AMIs mit einer Version 2018.03 zusammen mit dem Veröffentlichungsdatum zurückzugeben,

sodass Sie ein geeignetes Amazon Linux AMI als Basis auswählen können. MyRegion Ersetzen Sie es durch Ihre Regionskennung, z. B. us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

Version 5.16.0

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.16.0. Änderungen beziehen sich auf Version 5.15.0.

Erste Version: 19. Juli 2018

Upgrades

- 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
- JDBCRedshift-Treiber 1.2.15.1025

Änderungen, Verbesserungen und behobene Probleme

- HBase
 - [HBASERückportiert -20723](#)
- Presto
 - Konfigurationsänderungen zur Unterstützung der Authentifizierung. LDAP Weitere Informationen finden Sie unter [LDAPAuthentifizierung für Presto bei Amazon verwenden EMR](#).

- Spark
 - [Apache Spark Version 2.3.1, verfügbar ab EMR Amazon-Version 5.16.0, adressiert CVE-2018-8024 und -2018-1334. CVE](#) Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

Version 5.15.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.15.0. Änderungen beziehen sich auf Version 5.14.0.

Erste Version: 21. Juni 2018

Upgrades

- Auf 1.4.4 aktualisiert HBase
- Upgrade auf Hive 2.3.3
- Upgrade auf Hue 4.2.0
- Upgrade auf Oozie 5.0.0
- Upgrade auf ZooKeeper 3.4.12.
- Auf AWS SDK 1.11.333 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Hive
 - [Rückportiert -18069 HIVE](#)
- Hue
 - Hue wurde aktualisiert, um sich korrekt bei Livy zu authentifizieren, wenn Kerberos aktiviert ist. Livy wird jetzt unterstützt, wenn Kerberos mit Amazon verwendet wird. EMR
- JupyterHub
 - Aktualisiert, JupyterHub sodass Amazon standardmäßig LDAP Client-Bibliotheken EMR installiert.
 - Ein Fehler im Skript, das selbstsignierte Zertifikate generiert, wurde behoben.

Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

Version 5.14.1

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.14.1.

Änderungen beziehen sich auf Version 5.14.0.

Veröffentlichungsdatum: 17. Oktober 2018

Die Standardeinstellung AMI für Amazon wurde aktualisiertEMR, um potenzielle Sicherheitslücken zu beheben.

Version 5.14.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.14.0.

Änderungen beziehen sich auf Version 5.13.0.

Erste Version: 4. Juni 2018

Upgrades

- Upgrade auf Apache Flink 1.4.2
- Apache wurde auf 1.1.0 MXnet aktualisiert
- Apache Sqoop auf 1.4.7 aktualisiert

Neue Features

- JupyterHub Unterstützung hinzugefügt. Weitere Informationen finden Sie unter [JupyterHub](#).

Änderungen, Verbesserungen und behobene Probleme

- EMRFS
 - Die userAgent Zeichenfolge in Anfragen an Amazon S3 wurde aktualisiert und enthält nun die Benutzer- und Gruppeninformationen des aufrufenden Prinzipals. Dies kann zusammen mit AWS CloudTrail Protokollen für eine umfassendere Nachverfolgung von Anfragen verwendet werden.
- HBase
 - Enthalten ist [HBASE-20447](#), das ein Problem behebt, das zu Cache-Problemen führen könnte, insbesondere bei geteilten Regionen.
- MXnet
 - OpenCV-Plattformversion.
- Spark
 - Wenn Spark Parquet-Dateien mit an einen Amazon S3 S3-Speicherort schreibtEMRFS, wurde der FileOutputCommitter Algorithmus aktualisiert, sodass er Version 2 statt Version 1 verwendet. Dies reduziert die Anzahl der Umbenennungen, was die Anwendungsleistung verbessert. Diese Änderung beeinträchtigt nicht:
 - Andere Anwendungen als Spark.
 - Anwendungen, die in andere Dateisysteme schreiben, z. B. HDFS (die immer noch Version 1 von verwenden FileOutputCommitter).
 - Anwendungen, die andere Ausgabeformate wie Text oder CSV verwenden, die bereits EMRFS Direct Write verwenden.

Bekannte Probleme

- JupyterHub

- Die Verwendung von Konfigurationsklassifizierungen zur Einrichtung JupyterHub und individuellen Einrichtung von Jupyter-Notebooks beim Erstellen eines Clusters wird nicht unterstützt. Bearbeiten Sie die Dateien `jupyterhub_config.py` und `jupyter_notebook_config.py` für jeden Benutzer manuell. Weitere Informationen finden Sie unter [Konfiguration JupyterHub](#).
- JupyterHub kann auf Clustern innerhalb eines privaten Subnetzes nicht gestartet werden und schlägt mit der Meldung `feh1. Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` Dies wird durch einen Fehler im Skript verursacht, das selbstsignierte Zertifikate generiert. Verwenden Sie die folgende Problemumgehung zum Generieren selbstsignierter Zertifikate. Alle Befehle werden ausgeführt, während Sie mit dem Primärknoten verbunden sind.

1. Kopieren Sie das Skript zur Zertifikatsgenerierung aus dem Container auf den Primärknoten:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Verwenden Sie einen Texteditor, um Zeile 23 zu ändern, um den öffentlichen Hostnamen in den lokalen Hostnamen zu ändern, wie unten gezeigt:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Führen Sie das Skript aus, um selbstsignierte Zertifikate zu generieren:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Verschieben Sie die Zertifikatsdateien, die das Skript generiert, in das `/etc/jupyter/conf/`-Verzeichnis:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Sie können `tail` die `jupyter.log` Datei überprüfen, um zu überprüfen, ob sie JupyterHub neu gestartet wurde und einen 200-Antwortcode zurückgibt. Beispielsweise:

```
tail -f /var/log/jupyter/jupyter.log
```

Dadurch sollte eine Antwort ähnlich der folgenden ausgegeben:


```
# [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
```

- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in EMR Amazon-Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

Version 5.13.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.13.0. Änderungen beziehen sich auf Version 5.12.0.

Upgrades

- Upgrade auf Spark 2.3.0
- Auf 1.4.2 aktualisiert HBase
- Upgrade auf Presto 0.194
- Auf AWS SDK for Java 1.11.297 aktualisiert

Änderungen, Verbesserungen und behobene Probleme

- Hive
 - [Rückportiert -15436HIVE](#). Hive APIs wurde dahingehend verbessert, dass nur Ansichten zurückgegeben werden.

Bekannte Probleme

- MXNethat derzeit keine OpenCV-Bibliotheken.

Version 5.12.2

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.12.2. Änderungen beziehen sich auf Version 5.12.1

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

Version 5.12.1

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.12.1.

Änderungen beziehen sich auf Version 5.12.0.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiertEMR, um potenzielle Sicherheitslücken zu beheben.

Version 5.12.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.12.0.

Änderungen beziehen sich auf Version 5.11.1.

Upgrades

- AWS SDK für Java 1.11.238 ⇒ 1.11.267. Weitere Informationen finden Sie im Change Log on [AWS SDK für Java](#). GitHub
- Hadoop 2.7.3 ⇒ 2.8.3. Weitere Informationen finden Sie unter [Apache Hadoop im Releases](#).
- Flink 1.3.2 ⇒ 1.4.0. Weitere Informationen finden Sie in der [Versionsankündigung zu Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Weitere Informationen finden Sie in der [HBase Versionsankündigung](#).
- Hue 4.0.1 ⇒ 4.1.0. Weitere Informationen finden Sie in den [Versionshinweisen](#).
- MxNet 0.12.0 ⇒ 1.0.0. Weitere Informationen finden Sie im [MXNetChange](#) Log on. GitHub
- Presto 0.187 ⇒ 0.188. Weitere Informationen finden Sie in den [Versionshinweisen](#).

Änderungen, Verbesserungen und behobene Probleme

- Hadoop

- Die `yarn.resourcemanager.decommissioning.timeout`-Eigenschaft wurde zu `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs` geändert. Sie können diese Eigenschaft verwenden, um die Cluster-Skalierung anzupassen. Weitere Informationen finden Sie unter [Cluster Scale-Down](#) im Amazon EMR Management Guide.
- Hadoop CLI hat die `-d` Option zum Befehl `cp` (copy) hinzugefügt, der das direkte Kopieren spezifiziert. Sie können dies verwenden, um die Erstellung einer `.COPYING`-Zwischendatei zu vermeiden, wodurch das Kopieren von Daten zwischen Amazon S3 beschleunigt wird. Weitere Informationen finden Sie unter [HADOOP-12384](#).
- Pig
 - Die `pig-env`-Konfigurationsklassifizierung wurde hinzugefügt, die die Konfiguration der Pig-Umgebungseigenschaften vereinfacht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).
- Presto
 - Die Konfigurationsklassifizierung `presto-connector-redshift` wurde hinzugefügt, mit der Sie Werte in der Presto `redshift.properties` konfigurieren können. Weitere Informationen finden Sie unter [Redshift Connector](#) in der Presto-Dokumentation und [Anwendungen konfigurieren](#).
 - Presto-Unterstützung für EMRFS wurde hinzugefügt und ist die Standardkonfiguration. Frühere EMR Amazon-Versionen verwendeten `PrestoS3FileSystem`, was die einzige Option war. Weitere Informationen finden Sie unter [EMRFS und PrestoS3-Konfiguration FileSystem](#).

Note

Wenn Sie die zugrunde liegenden Daten in Amazon S3 mit Amazon EMR Version 5.12.0 abfragen, können Presto-Fehler auftreten. Dies liegt daran, dass Presto die Konfigurationsklassifizierungswerte von `emrfs-site.xml` nicht abrufen. Um dieses Problem zu umgehen, erstellen Sie ein `emrfs`-Unterverzeichnis unter `usr/lib/presto/plugin/hive-hadoop2/` und erstellen einen Symlink `usr/lib/presto/plugin/hive-hadoop2/emrfs` zu der vorhandenen Datei `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml`. Starten Sie dann den Presto-Server-Prozess neu (`sudo presto-server stop` gefolgt von `sudo presto-server start`).

- Spark
 - Backported [SPARK-22036](#): Die Multiplikation gibt manchmal Null zurück. `BigDecimal`

Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- SparkR ist nicht für Cluster verfügbar, die mit einem benutzerdefinierten System erstellt wurden AMI, da R nicht standardmäßig auf Clusterknoten installiert ist.

Version 5.11.3

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.11.3. Änderungen beziehen sich auf Version 5.11.2.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Die Standardversion von Amazon Linux AMI für Amazon wurde aktualisiert und enthält EMR nun wichtige Sicherheitsupdates für den Linux-Kernel, einschließlich des TCP SACK Denial-of-Service-Problems ([AWS-2019-005](#)).

Version 5.11.2

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.11.2. Änderungen beziehen sich auf Version 5.11.1.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

Version 5.11.1

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.11.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.11.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Bekannte Probleme

- MXNetenthält keine OpenCV-Bibliotheken.
- Hive 2.3.2 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

Version 5.11.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.11.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.10.0.

Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen.

- Hive 2.3.2
- Spark 2.2.1
- SDK für Java 1.11.238

Neue Features

- Spark
 - Die Einstellung `spark.decommissioning.timeout.threshold` wurde hinzugefügt, was das Verhalten von Spark bei der Außerbetriebnahme bei Verwendung von Spot-Instances verbessert. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
 - Die `aws-sagemaker-spark-sdk` Komponente wurde zu Spark hinzugefügt, die Amazon SageMaker Spark und zugehörige Abhängigkeiten für die Spark-Integration mit [Amazon](#) installiert SageMaker. Sie können Amazon SageMaker Spark verwenden, um Spark-Pipelines für maschinelles Lernen (ML) mithilfe von SageMaker Amazon-Stufen zu erstellen. Weitere Informationen finden Sie in der [SageMaker Spark-Readme-Datei](#) GitHub und unter [Using Apache Spark with Amazon SageMaker](#) im Amazon SageMaker Developer Guide.

Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.2 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

Version 5.10.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.10.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.9.0.

Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen.

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

Neue Features

- Die Kerberos-Authentifizierung wird jetzt unterstützt. Weitere Informationen finden Sie unter [Verwenden der Kerberos-Authentifizierung](#) im Amazon Management Guide EMR
- Unterstützung für IAM Rollen für EMRFS Anfragen an Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter [IAM Rollen für EMRFS Anfragen an Amazon S3 konfigurieren](#) im Amazon EMR Management Guide.
- Unterstützung für GPU basierte P2- und P3-Instance-Typen hinzugefügt. Weitere Informationen finden Sie unter [Amazon EC2 P2-Instances](#) und [Amazon EC2 P3-Instances](#). NVIDIA-Treiber 384.81 und CUDA Treiber 9.0.176 sind standardmäßig auf diesen Instance-Typen installiert.
- Unterstützung für [Apache MXNet](#) hinzugefügt.

Änderungen, Verbesserungen und behobene Probleme

- Presto
 - Unterstützung für die Verwendung des AWS Glue-Datenkatalogs als Standard-Hive-Metastore hinzugefügt. Weitere Informationen finden Sie unter [Presto mit dem AWS Glue-Datenkatalog verwenden](#).
 - Unterstützung von [koordinatenbasierten Funktionen](#) hinzugefügt.
 - Unterstützung eines [Überlaufs auf die Festplatte](#) für Joins hinzugefügt.
 - Unterstützung für den [Redshift-Connector](#) hinzugefügt.
- Spark
 - [SPARK-20640](#) wurde zurückportiert, wodurch das RPC-Timeout und die Wiederholungsversuche für Shuffle-Registrierungswerte mithilfe von Eigenschaften konfiguriert werden können. `spark.shuffle.registration.timeout`
`spark.shuffle.registration.maxAttempts`
 - [SPARK-21549](#) wurde zurückportiert, wodurch ein Fehler behoben wird, der beim Schreiben von benutzerdefinierten Daten an andere Speicherorte auftritt. OutputFormat HDFS
- Rückportierung von [Hadoop-13270](#)

- Die Bibliotheken Numpy, Scipy und Matplotlib wurden aus dem Basis-Arizona entfernt. EMR AMI Wenn diese Bibliotheken für Ihre Anwendung erforderlich sind, werden sie im folgenden Repository bereitgestellt, sodass Sie sie unter Verwendung einer Bootstrap-Aktion mit `yum install` auf allen Knoten installieren können.
- In der EMR Amazon-Basis sind AMI keine RPM Anwendungspakete mehr enthalten, sodass die RPM Pakete nicht mehr auf Clusterknoten vorhanden sind. Custom AMIs und Amazon EMR Base verweisen AMI jetzt auf das RPM Paket-Repository in Amazon S3.
- Aufgrund der Einführung der sekundengenauen Abrechnung in Amazon EC2 lautet das Standardverhalten beim Herunterfahren jetzt bei Abschluss der Aufgabe beenden und nicht mehr bei Instance-Stunde beenden. Weitere Informationen finden Sie unter [Konfigurieren eines Cluster-Scale-Downs](#).

Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.1 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

Version 5.9.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.9.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.8.0.

Veröffentlichungsdatum: 5. Oktober 2017

Letzte Aktualisierung der Funktionen: 12. Oktober 2017

Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen.

- AWS SDK for Java Version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Neue Features

- Livy-Support hinzugefügt (Version 0.4.0-incubating). Weitere Informationen finden Sie unter [Apache Livy](#).
- Support für Hue-Notebooks für Spark hinzugefügt.
- Unterstützung für EC2 Amazon-Instances der i3-Serie hinzugefügt (12. Oktober 2017).

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Neue Funktionen hinzugefügt, die sicherstellen, dass Spark die Knotenbeendigung infolge einer manuellen Größenänderung oder einer Anforderung einer Auto Scaling-Richtlinie zuverlässiger ausführt. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
 - SSL wird anstelle von 3 DES für die Verschlüsselung während der Übertragung für den Blocktransfer-Service verwendet, wodurch die Leistung bei der Verwendung von EC2 Amazon-Instance-Typen mit AES -NI verbessert wird.
 - [Rückportiert -21494SPARK](#).
- Zeppelin
 - [Rückportiert -2377. ZEPPELIN](#)
- HBase
 - Patch [HBASE-18533](#) wurde hinzugefügt, der zusätzliche Werte für die Konfiguration mithilfe der HBase BucketCache Konfigurationsklassifizierung ermöglicht. `hbase-site`

- Hue
 - AWS Glue Data Catalog-Unterstützung für den Hive-Abfrageeditor in Hue hinzugefügt.
 - Standardmäßig können Superuser in Hue auf alle Dateien zugreifen, auf die EMR IAM Amazon-Rollen zugreifen dürfen. Neu erstellte Benutzer verfügen nicht automatisch über Zugriffsberechtigungen auf den Amazon-S3-Filebrowser und für ihre Gruppen müssen die `filebrowser.s3_access`-Berechtigungen aktiviert sein.
- Es wurde ein Problem behoben, das dazu führte, dass auf die zugrunde liegenden JSON Daten, die mit dem AWS Glue-Datenkatalog erstellt wurden, nicht zugegriffen werden konnte.

Bekannte Probleme

- Der Clusterstart schlägt fehl, wenn alle Anwendungen installiert sind und die Standardgröße des EBS Amazon-Root-Volumes nicht geändert wird. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren `--ebs-root-volume-size` Parameter an.
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle `LOCATION` hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

Version 5.8.2

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.8.2. Änderungen beziehen sich auf Version 5.8.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert EMR, um potenzielle Sicherheitslücken zu beheben.

Version 5.8.1

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.8.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.8.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Version 5.8.0

Die folgenden Versionshinweise enthalten Informationen zur EMR Amazon-Version 5.8.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.7.0.

Erste Version: 10. August 2017

Letzte Aktualisierung der Funktionen: 25. September 2017

Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen:

- AWS SDK1.11.160
- Flink 1.3.1
- Hive 2.3.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache-Hive-Website.
- Spark 2.2.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache-Spark-Website.

Neue Features

- Support zur Anzeige des Anwendungsverlaufs hinzugefügt (25. September 2017). Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Amazon EMR Management Guide.

Änderungen, Verbesserungen und behobene Probleme

- Integration mit dem AWS Glue-Datenkatalog
 - Hive und Spark haben jetzt die Möglichkeit SQL, AWS Glue Data Catalog als Hive-Metadatenpeicher zu verwenden. Weitere Informationen erhalten Sie unter [Den AWS Glue-Datenkatalog als Metastore für Hive verwenden](#) und [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).
- Der Anwendungsverlauf wurde zu den Cluster-Details hinzugefügt, sodass Sie historische Daten für YARN Anwendungen und zusätzliche Details für Spark-Anwendungen anzeigen können. Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Amazon EMR Management Guide.
- Oozie
 - Rückportiert [OOZIE-2748](#).
- Hue
 - Rückportiert [HUE-5859](#)
- HBase
 - Es wurde ein Patch hinzugefügt, um die Startzeit des HBase Masterservers mithilfe von Java Management Extensions (JMX) anzuzeigen. `getMasterInitializedTime`
 - Patch zur Verbesserung der Cluster-Startzeit hinzugefügt.

Bekanntete Probleme

- Der Clusterstart schlägt fehl, wenn alle Anwendungen installiert sind und die Standardgröße des EBS Amazon-Root-Volumes nicht geändert wird. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren `--ebs-root-volume-size` Parameter an.
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

- Spark – Bei der Verwendung von Spark tritt bei dem appusher-Daemon bei Spark-Aufträgen mit langer Ausführungsdauer nach einigen Stunden oder Tagen möglicherweise ein Speicherleck-Problem bei dem Datei-Handler auf. Um das Problem zu beheben, stellen Sie eine Verbindung mit dem Master-Knoten her und geben `sudo /etc/init.d/appusher stop` ein. Dadurch wird der Appusher-Daemon gestoppt, den Amazon EMR automatisch neu startet.
- Application history
 - Es sind keine Verlaufsdaten für beendete Spark-Executors verfügbar.
 - Der Anwendungsverlauf ist nicht für Cluster verfügbar, die eine Sicherheitskonfiguration verwenden, um die Verschlüsselung bei der Übertragung zu aktivieren.

Version 5.7.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.7.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.6.0.

Veröffentlichungsdatum: 13. Juli 2017

Upgrades

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Neue Features

- Es wurde die Möglichkeit hinzugefügt, ein benutzerdefiniertes Amazon Linux anzugebenAMI, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Benutzerdefiniert verwenden AMI](#).

Änderungen, Verbesserungen und behobene Probleme

- HBase

- Es wurde die Möglichkeit hinzugefügt, HBase Read-Replica-Cluster zu konfigurieren. Weitere Informationen finden Sie unter [Verwenden eines Read-Replica-Clusters](#).
- Mehrere Fehlerbehebungen und Verbesserungen
- Presto – neue Konfigurationsfunktion `node.properties` hinzugefügt.
- YARN- Fähigkeit zur Konfiguration hinzugefügt `container-log4j.properties`
- Sqoop - rückportiert [SQOOP-2880](#), was ein Argument einführt, mit dem Sie das temporäre Sqoop-Verzeichnis festlegen können.

Version 5.6.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.6.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.5.0.

Veröffentlichungsdatum: 5. Juni 2017

Upgrades

- Flink 1.2.1
- HBase 1.3.1
- Mahout 0.13.0. Dies ist die erste Version von Mahout, die Spark 2.x in EMR Amazon-Version 5.0 und höher unterstützt.
- Spark 2.1.1

Änderungen, Verbesserungen und behobene Probleme

- Presto
 - Es wurde die Möglichkeit hinzugefügt, die Kommunikation zwischen Presto-Knoten zu SSL aktivieren/zu TLS sichern, indem die Verschlüsselung während der Übertragung mithilfe einer Sicherheitskonfiguration aktiviert wird. Weitere Informationen finden Sie unter [Datenverschlüsselung während der Übertragung](#).
 - [Presto 7661](#) wurde rückportiert, wodurch der Anweisung `EXPLAIN ANALYZE` die Option `VERBOSE` hinzugefügt wurde, um detailliertere Low-Level-Statistiken über einen Abfrageplan bereitzustellen.

Version 5.5.3

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.5.3. Änderungen beziehen sich auf Version 5.5.2.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

Version 5.5.2

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 5.5.2. Änderungen beziehen sich auf Version 5.5.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiertEMR, um potenzielle Sicherheitslücken zu beheben.

Version 5.5.1

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.5.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.5.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von defaultAmazon Linux AMI for Amazon wurde aktualisiert, EMR um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Version 5.5.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.5.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.4.0.

Veröffentlichungsdatum: 26. April 2017

Upgrades

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Änderungen, Verbesserungen und behobene Probleme

- Spark
 - Der Spark-Patch ([SPARK-20115](#)) wurde zurückportiert, [DAGScheduler um alle verlorenen Shuffle-Blöcke neu zu berechnen, wenn der externe Shuffle-Dienst für Version 2.1.0 von Spark, die in dieser Version enthalten ist, nicht verfügbar ist.](#)
- Flink
 - Flink ist nun in Scala 2.11 integriert. Wenn Sie Scala API und Bibliotheken verwenden, empfehlen wir Ihnen, Scala 2.11 in Ihren Projekten zu verwenden.
 - Ein Problem wurde behoben, bei dem die Standardeinstellungen `HADOOP_CONF_DIR` und `YARN_CONF_DIR` nicht ordnungsgemäß festgelegt wurden, sodass `start-scala-shell.sh` nicht funktionierte. Außerdem wurde die Möglichkeit hinzugefügt, diese Werte mit `env.hadoop.conf.dir` und `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` oder der Konfigurationsklassifizierung `flink-conf` festzulegen.
 - Es wurde ein neuer EMR -spezifischer Befehl `flink-scala-shell` als Wrapper für eingeführt. `start-scala-shell.sh` Wir empfehlen, diese Befehl statt `start-scala-shell` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-scala-shell -n 2` eine Flink Scala-Shell mit einer Aufgabenparallelität von 2.
 - Es wurde ein neuer EMR -spezifischer Befehl `flink-yarn-session` als Wrapper für eingeführt. `yarn-session.sh` Wir empfehlen, diese Befehl statt `yarn-session` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-`

`yarn-session -d -n 2` eine langlebige Flink-Sitzung in einem getrennten Status mit zwei Aufgabenmanagern.

- Der adressierte [\(FLINK-6125\) Commons-HttpClient ist in Flink 1.2 nicht mehr schattiert](#).
- Presto
 - Unterstützung für Authentifizierung hinzugefügt. LDAP Für die Verwendung LDAP mit Presto bei Amazon EMR müssen Sie den HTTPS Zugriff für den Presto-Koordinator (`http-server.https.enabled=true` in `config.properties`) aktivieren. Einzelheiten zur Konfiguration finden Sie in der [LDAPPresto-Dokumentation unter Authentifizierung](#).
 - Unterstützung für `SHOW GRANTS` hinzugefügt.
- Amazon EMR Base Linux AMI
 - EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie in den [Versionshinweisen zu Amazon Linux AMI 2017.03](#).
 - Python 2.6 wurde aus dem EMR Linux-Basisimage von Amazon entfernt. Python 2.7 und 3.4 sind standardmäßig installiert. Sie können Python 2.6 bei Bedarf manuell installieren.

Version 5.4.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.4.0. Die Änderungen beziehen sich auf die Version EMR 5.3.0 von Amazon.

Veröffentlichungsdatum: 8. März 2017

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Flink 1.2.0
- Upgrade auf Hbase 1.3.0
- Upgrade auf Phoenix 4.9.0

Note

Wenn Sie ein Upgrade von einer früheren Version von Amazon EMR auf EMR Amazon-Version 5.4.0 oder höher durchführen und die sekundäre Indizierung verwenden, aktualisieren Sie lokale Indizes wie in der [Apache Phoenix-Dokumentation](#) beschrieben. Amazon EMR entfernt die erforderlichen Konfigurationen aus der `hbase-site`

Klassifizierung, aber die Indizes müssen neu gefüllt werden. Online- und Offline-Upgrades von Indizes werden unterstützt. Online-Upgrades sind die Standardeinstellung. Dies bedeutet, dass Indizes neu gefüllt werden, während die Initialisierung von Phoenix-Clients Version 4.8.0 oder höher ausgeführt wird. Um Offline-Upgrades anzugeben, setzen Sie die `phoenix.client.localIndexUpgrade` Konfiguration in der `phoenix-site` Klassifizierung auf `False` und dann SSH auf den Master-Knoten, der ausgeführt werden soll. `psql [zookeeper] -1`

- Upgrade auf Presto 0.166
- Upgrade auf Zeppelin 0.7.0

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label `emr-5.4.0` vorgenommen:

- Es werden jetzt `r4`-Instances unterstützt. Siehe [EC2 Amazon-Instance-Typen](#).

Version 5.3.1

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.3.1. Die Änderungen beziehen sich auf die Version EMR 5.3.0 von Amazon.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen am Backport von Zeppelin-Patches und an der Aktualisierung der Standardeinstellung AMI für Amazon. EMR

Version 5.3.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.3.0. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.2.1.

Veröffentlichungsdatum: 26. Januar 2017

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hive 2.1.1
- Upgrade auf Hue 3.11.0
- Upgrade auf Spark 2.1.0
- Upgrade auf Oozie 4.3.0
- Upgrade auf Flink 1.1.4

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label emr-5.3.0 vorgenommen:

- Ein Patch für Hue wurde hinzugefügt, mit dem Sie die Einstellung `interpreters_shown_on_wheel` verwenden können, um zu konfigurieren, was Interpreter in der Notebook-Auswahl als Erstes anzeigen, unabhängig von ihrer Reihenfolge in der Datei `hue.ini`.
- Die Konfigurationsklassifizierung `hive-parquet-logging` wurde hinzugefügt, mit der Sie Werte in der Hive-Datei `parquet-logging.properties` konfigurieren können.

Version 5.2.2

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.2.2. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.2.1.

Veröffentlichungsdatum: 2. Mai 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Backported [SPARK-194459](#), wodurch ein Problem behoben wurde, bei dem das Lesen aus einer ORC Tabelle mit `char`/`varchar`-Spalten fehlschlagen kann.

Version 5.2.1

Die folgenden Versionshinweise enthalten Informationen zur Amazon-Version EMR 5.2.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 5.2.0.

Veröffentlichungsdatum: 29. Dezember 2016

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf ZooKeeper 3.4.9. Weitere Informationen finden Sie in den [ZooKeeper Versionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label emr-5.2.1 vorgenommen:

- Unterstützung für den Instance-Typ Amazon EC2 m4.16xlarge in EMR Amazon-Version 4.8.3 und höher hinzugefügt, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- Der Speicherort von Flink und die YARN Konfigurationspfade sind jetzt standardmäßig festgelegt/etc/default/flink, sodass Sie die Umgebungsvariablen nicht festlegen müssen FLINK_CONF_DIR und HADOOP_CONF_DIR wenn Sie die `flink` oder `yarn-session.sh` - Treiberskripte ausführen, um Flink-Jobs zu starten.
- Unterstützung für FlinkKinesisConsumer Klassen hinzugefügt.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem ControlledJob # mit einer Nullzeiger-Ausnahme (NPE) toString fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

Version 5.2.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.2.0. Die Änderungen beziehen sich auf die Version EMR 5.1.0 von Amazon.

Veröffentlichungsdatum: 21. November 2016

Änderungen und Verbesserungen

Die folgenden Änderungen und Verbesserungen sind in dieser Version verfügbar:

- Amazon S3 S3-Speichermodus für hinzugefügtHBase.
- Ermöglicht es Ihnen, einen Amazon S3 S3-Speicherort für das HBase Rootdir anzugeben. Weitere Informationen finden Sie [HBaseunter Amazon S3](#).

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Spark 2.0.2

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem /mnt bei Instance-Typen EBS nur auf 2 TB beschränkt war.
- Es wurde ein Problem mit Instance-Controller und Logpusher-Protokollen behoben, die an die entsprechenden OUT-Dateien anstelle ihrer üblichen log4j-konfigurierten LOG-Dateien, die stündlich rotieren, ausgegeben wurden. Da die OUT-Dateien nicht rotieren, ist letztendlich die /emr-Partition belegt. Dieses Problem betrifft nur Hardware-Instanztypen für virtuelle Maschinen (HVM).

Version 5.1.0

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.1.0. Die Änderungen beziehen sich auf die Version EMR 5.0.0 von Amazon.

Veröffentlichungsdatum: 03. November 2016

Änderungen und Verbesserungen

Die folgenden Änderungen und Verbesserungen sind in dieser Version verfügbar:

- Flink 1.1.3 wird jetzt unterstützt.
- Presto wurde als Option im Notebookabschnitt von Hue hinzugefügt.

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Auf 1.2.3 aktualisiert HBase
- Upgrade auf Zeppelin 0.6.2

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem mit Tez-Abfragen auf Amazon S3 behoben, bei ORC denen Dateien nicht so gut funktionierten wie frühere Amazon EMR 4.x-Versionen.

Version 5.0.3

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 5.0.3. Die Änderungen beziehen sich auf die Version EMR 5.0.0 von Amazon.

Veröffentlichungsdatum: 24. Oktober 2016

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hadoop 2.7.3
- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Webschnittstelle](#) in der Presto-Dokumentation.
- Upgrade auf Spark 2.0.1
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Version 5.0.0

Veröffentlichungsdatum: 27. Juli 2016

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hive 2.1
- Upgrade auf Presto 0.150
- Upgrade auf Spark 2.0
- Upgrade auf Hue 3.10.0
- Upgrade auf Pig 0.16.0
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin 0.6.1

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label emr-5.0.0 oder höher vorgenommen:

- Amazon EMR unterstützt die neuesten Open-Source-Versionen von Hive (Version 2.1) und Pig (Version 0.16.0). Wenn Sie Hive oder Pig EMR in der Vergangenheit bei Amazon verwendet haben, kann sich dies auf einige Anwendungsfälle auswirken. Weitere Informationen finden Sie unter [Hive](#) und [Pig](#).
- Die Standard-Ausführungs-Engine für Hive und Pig ist jetzt Tez. Um dies zu ändern, bearbeiten Sie die entsprechenden Werte in den Konfigurationsklassifizierungen `hive-site` bzw. `pig-properties`.
- Eine verbesserte Schritt-Debugging-Funktion wurde hinzugefügt. Mit dieser Funktion können Sie die Ursache von Schrittfehlern sehen, wenn der Service die Ursache ermitteln kann. Weitere Informationen finden Sie unter [Enhanced Step Debugging](#) im Amazon EMR Management Guide.
- Anwendungen, die zuvor mit "-Sandbox" endeten, tragen diesen Suffix nicht mehr. Dies kann Ihre Automatisierung, zum Beispiel, wenn Sie Skripts zum Starten von Clustern mit diesen Anwendungen verwenden, unterbrechen. Die folgende Tabelle zeigt Anwendungsnamen in Amazon EMR 4.7.2 im Vergleich zu Amazon EMR 5.0.0.

Änderungen der Anwendungsnamen

Amazon EMR 4,7,2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Sandkasten	ZooKeeper

- Spark ist jetzt für Scala 2.11 kompiliert.
- Java 8 ist jetzt der Standard. JVM Alle Anwendungen werden mit der Java 8-Laufzeit ausgeführt. Es sind keine Änderungen am Byte-Codeziel der Anwendungen vorgenommen worden. Die meisten Anwendungen zielen weiterhin auf Java 7 ab.
- Zeppelin enthält jetzt Authentifizierungsfunktionen. Weitere Informationen finden Sie unter [Zeppelin](#).
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

Version 4.9.5

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 4.9.5.

Änderungen beziehen sich auf Version 4.9.4.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- HBase
 - Diese Version behebt eine potenzielle Schwachstellen.

Version 4.9.4

Die folgenden Versionshinweise enthalten Informationen für EMR Amazon-Version 4.9.4. Änderungen beziehen sich auf Version 4.9.3.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert, um potenzielle Sicherheitslücken zu beheben.

Version 4.9.3

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.9.3. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.9.2.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon Linux-Kernel von default Amazon Linux AMI for Amazon wurde aktualisiert, um Sicherheitslücken im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, -2017-5753 und CVE -2017-5754). CVE Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Version 4.9.2

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.9.2. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.9.1.

Veröffentlichungsdatum: 13. Juli 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

Version 4.9.1

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.9.1. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.4.

Veröffentlichungsdatum: 10. April 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- [Backports von HIVE-9976 und -10106 HIVE](#)
- Es wurde ein Problem behoben YARN, bei dem eine große Anzahl von Knoten (mehr als 2.000) und Containern (mehr als 5.000) zu einem Fehler aufgrund unzureichenden Speichers führte, zum Beispiel: "Exception in thread 'main' java.lang.OutOfMemoryError"

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label emr-4.9.1 vorgenommen:

- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- Python 2.6 wurde aus dem EMR Linux-Basisimage von Amazon entfernt. Sie können Python 2.6 bei Bedarf manuell installieren.

Version 4.8.4

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.4. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.3.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

Version 4.8.3

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.3. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.2.

Veröffentlichungsdatum: 29. Dezember 2016

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf Spark 1.6.3. Weitere Informationen finden Sie in den [Spark-Versionshinweisen](#) in der Apache-Spark-Dokumentation.
- Auf ZooKeeper 3.4.9 aktualisiert. Weitere Informationen finden Sie in den [ZooKeeper Versionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label emr-4.8.3 vorgenommen:

- Unterstützung für den Instance-Typ Amazon EC2 m4.16xlarge in EMR Amazon-Version 4.8.3 und höher hinzugefügt, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem ControlledJob # mit einer Nullzeiger-Ausnahme (NPE) toString fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

Version 4.8.2

Die folgenden Versionshinweise enthalten Informationen für die Amazon-Version EMR 4.8.2. Die Änderungen beziehen sich auf die Amazon-Version EMR 4.8.0.

Veröffentlichungsdatum: 24. Oktober 2016

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hadoop 2.7.3

- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Webschnittstelle](#) in der Presto-Dokumentation.
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Version 4.8.0

Veröffentlichungsdatum: 7. September 2016

Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Auf 1.2.2 aktualisiert HBase
- Upgrade auf Presto-Sandbox 0.151
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin-Sandbox 0.6.1

Änderungen und Verbesserungen

Die folgenden Änderungen wurden an EMR Amazon-Versionen für das Release-Label emr-4.8.0 vorgenommen:

- Es wurde ein Problem behoben YARN, bei dem versucht ApplicationMaster wurde, Container zu bereinigen, die nicht mehr existieren, weil ihre Instances beendet wurden.
- Die Aktionen hive-server2 URL für Hive2 in den Oozie-Beispielen wurden korrigiert.
- Zusätzliche Presto-Kataloge werden jetzt unterstützt.
- [Rückportierte Patches: HIVE-8948, -12679, -13405, -3116, -12689 HIVE HIVE PHOENIX HADOOP](#)
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

Version 4.7.2

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR 4.7.2.

Veröffentlichungsdatum: 15. Juli 2016

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Upgrade auf Mahout 0.12.2
- Upgrade auf Presto 0.148
- Upgrade auf Spark 1.6.2
- Sie können jetzt eine `AWSCredentialsProvider` für die EMRFS Verwendung mit a URI als Parameter erstellen. Weitere Informationen finden Sie unter [Erstellen eines `AWSCredentialsProvider` für EMRFS](#).
- EMRFS ermöglicht es Benutzern jetzt, mithilfe der `fs.s3.consistent.dynamodb.endpoint` Eigenschaft in einen benutzerdefinierten DynamoDB-Endpunkt für ihre Consistent View-Metadaten zu konfigurieren. `emrfs-site.xml`
- Es wurde ein Skript in `/usr/bin` mit dem Namen `spark-example` hinzugefügt, das `/usr/lib/spark/spark/bin/run-example` umschließt, sodass Sie Beispiele direkt ausführen können. Um beispielsweise das SparkPi Beispiel auszuführen, das in der Spark-Distribution enthalten ist, können Sie es `spark-example SparkPi 100` von der Befehlszeile aus ausführen oder `command-runner.jar` als Schritt in der API

Bekanntes Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem Oozie die Datei `spark-assembly.jar` nicht am richtigen Speicherort ablegte, wenn Spark auch installiert war. Dies führte dazu, dass Spark-Anwendungen nicht mit Oozie gestartet werden konnten.
- Ein Problem mit der auf Spark Log4J basierenden Protokollierung in Containern wurde behoben.
YARN

Version 4.7.1

Veröffentlichungsdatum: 10. Juni 2016

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, das die Startzeit von Clustern verlängerte, die in einem VPC mit privaten Subnetzen gestartet wurden. Der Fehler betraf nur Cluster, die mit der EMR Amazon-Version 4.7.0 gestartet wurden.
- Es wurde ein Problem behoben, bei dem das Auflisten von Dateien in Amazon EMR für Cluster, die mit der Version Amazon EMR 4.7.0 gestartet wurden, nicht richtig behandelt wurde.

Version 4.7.0

Important

Amazon EMR 4.7.0 ist veraltet. Verwenden Sie stattdessen Amazon EMR 4.7.1 oder höher.

Veröffentlichungsdatum: 2. Juni 2016

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Apache Phoenix 4.7.0 wurde hinzugefügt.
- Apache Tez 0.8.3 wurde hinzugefügt.
- Auf 1.2.1 aktualisiert HBase
- Upgrade auf Mahout 0.12.0
- Upgrade auf Presto 0.147
- Auf 1.10.75 AWS SDK for Java aktualisiert
- Das letzte Flag wurde von der Eigenschaft `mapreduce.cluster.local.dir` in `mapred-site.xml` entfernt, damit Benutzer Pig im lokalen Modus ausführen können.

Amazon Redshift JDBC Redshift-Treiber im Cluster verfügbar

Amazon Redshift JDBC Redshift-Treiber sind jetzt unter `/usr/share/aws/redshift/jdbc` enthalten. `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` ist der JDBC 4.1-kompatible Amazon Redshift Redshift-Treiber und `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` ist der JDBC 4.0-kompatible Amazon Redshift Redshift-Treiber. Weitere

Informationen finden [Sie unter JDBC Verbindung konfigurieren](#) im Amazon Redshift Management Guide.

Java 8

Mit Ausnahme von Presto ist Open JDK 1.7 der Standard, der für alle JDK Anwendungen verwendet wird. Sowohl Open JDK 1.7 als auch 1.8 sind jedoch installiert. Informationen zum Festlegen von JAVA_HOME für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Kernelproblem behoben, das die Leistung von durchsatzoptimierten HDD (st1) EBS Volumes für Amazon EMR in emr-4.6.0 erheblich beeinträchtigte.
- Es wurde ein Problem behoben, bei dem ein Cluster fehlschlug, wenn eine HDFS Verschlüsselungszone angegeben wurde, ohne Hadoop als Anwendung auszuwählen.
- Die HDFS Standard-Schreibrichtlinie wurde von RoundRobin zu AvailableSpaceVolumeChoosingPolicy geändert. Einige Volumes wurden mit der RoundRobin Konfiguration nicht richtig genutzt, was zu einem Ausfall der Kernknoten und zu unzuverlässigen HDFS Verbindungen führte.
- Es wurde ein Problem mit dem behoben EMRFSCLI, das bei der Erstellung der standardmäßigen DynamoDB-Metadatentabelle für konsistente Ansichten zu einer Ausnahme führte.
- Es wurde ein Deadlock-Problem behobenEMRFS, das möglicherweise bei mehrteiligen Umbenennungs- und Kopiervorgängen auftrat.
- Es wurde ein Problem behobenEMRFS, das dazu führte, dass die CopyPart Standardgröße 5 MB betrug. Der Standard ist jetzt ordnungsgemäß auf 128 MB eingestellt.
- Es wurde ein Problem mit der Zeppelin-Startkonfiguration behoben, das möglicherweise das Anhalten des Service verhinderte.
- Es wurde ein Problem mit Spark und Zeppelin behoben, das Sie daran hinderte, das s3a:// URI Schema zu verwenden, weil es nicht richtig in den jeweiligen Klassenpfad geladen /usr/lib/hadoop/hadoop-aws.jar wurde.
- [Rückportiert -2484. HUE](#)
- Ein [Commit](#) von Hue 3.9.0 (JIRA existiert nicht) wurde zurückportiert, um ein Problem mit dem Browser-Beispiel zu beheben. HBase
- [Rückportiert -9073HIVE](#).

Version 4.6.0

Veröffentlichungsdatum: 21. April 2016

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- HBase1.2.0 hinzugefügt
- ZooKeeper-Sandbox 3.4.8 wurde hinzugefügt.
- Upgrade auf Presto-Sandbox 0.143
- EMRAmazon-Versionen basieren jetzt auf Amazon Linux 2016.03.0. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.

Problem mit Auswirkungen auf durchsatzoptimierte HDD (st1) Volumetypen EBS

Ein Problem in den Linux-Kernel-Versionen 4.2 und höher beeinträchtigt die Leistung von durchsatzoptimierten HDD (st1) EBS Volumes für erheblich. EMR Diese Version (emr-4.6.0) verwendet die Kernel-Version 4.4.5 und ist daher betroffen. Daher empfehlen wir, emr-4.6.0 nicht zu verwenden, wenn Sie st1-Volumes verwenden möchten. EBS Sie können emr-4.5.0 oder frühere EMR Amazon-Versionen mit st1 ohne Auswirkungen verwenden. Außerdem stellen wir die Problembeseitigung mit späteren Versionen zur Verfügung.

Python-Standard Einstellungen

Python 3.4 ist jetzt standardmäßig installiert, Python 2.7 bleibt jedoch der Systemstandard. Sie können Python 3.4 mit einer Bootstrap-Aktion als Systemstandard konfigurieren. Sie können die Konfiguration verwenden, API um `PYSPARK_PYTHON` export `/usr/bin/python3.4` in der `spark-env` Klassifizierung festzulegen, um die Python-Version zu beeinflussen, die von PySpark verwendet wird.

Java 8

Mit Ausnahme von Presto ist Open JDK 1.7 der Standard, der für alle JDK Anwendungen verwendet wird. Sowohl Open JDK 1.7 als auch 1.8 sind jedoch installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die Anwendungsbereitstellung manchmal wegen eines generierten Passwort nach dem Zufallsprinzip ausfiel.
- Bisher war `mysqld` auf allen Knoten installiert. Jetzt ist es nur auf der Master-Instance installiert und nur dann, wenn die gewählte Anwendung `mysql-server` als Komponente enthält. Derzeit enthalten die folgenden Anwendungen die `mysql-server` Komponente: Hive, HueHCatalog, Presto-Sandbox und Sqoop-Sandbox.
- Von der Standardeinstellung `32 yarn.scheduler.maximum-allocation-vcores` auf 80 geändert, wodurch ein in `emr-4.4.0` eingeführtes Problem behoben wird, das hauptsächlich bei Spark auftritt, wenn die `maximizeResourceAllocation` Option in einem Cluster verwendet wird, dessen Core-Instance-Typ einer von wenigen großen Instance-Typen ist, bei denen die YARN `vcores` höher als 32 eingestellt sind; nämlich `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` oder `m4.10xlarge` wurden von diesem Problem betroffen.
- `s3-dist-cp` verwendet jetzt EMRFS für alle Amazon S3 S3-Nominierungen und speichert nicht mehr in einem temporären Verzeichnis. HDFS
- Es wurde ein Problem mit der Ausnahmebehandlung für die clientseitigen Verschlüsselung mehrteiliger Uploads behoben.
- Es wurde eine Option hinzugefügt, mit der Benutzer die Amazon-S3-Speicherklassen ändern können. Standardmäßig ist diese Einstellung `STANDARD`. Die Konfigurationsklassifizierungseinstellung `emrfs-site` ist `fs.s3.storageClass` und die möglichen Werte sind `STANDARD`, `STANDARD_IA` und `REDUCED_REDUNDANCY`. Weitere Informationen zu Speicherklassen finden Sie unter [Speicherklassen](#) im Benutzerhandbuch zu Amazon Simple Storage Service.

Version 4.5.0

Veröffentlichungsdatum: 4. April 2016

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Upgrade auf Spark 1.6.1
- Upgrade auf Hadoop 2.7.2

- Upgrade auf Presto 0.140
- AWS KMS Unterstützung für serverseitige Amazon S3 S3-Verschlüsselung hinzugefügt.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem My SQL - und Apache-Server nach dem Neustart eines Knotens nicht gestartet wurden.
- Es wurde ein Problem behoben, das mit IMPORT nicht partitionierten Tabellen, die in Amazon S3 gespeichert waren, nicht richtig funktionierte
- Es wurde ein Problem mit Presto behoben, bei dem das Staging-Verzeichnis beim Schreiben in Hive-Tabellen /mnt/tmp anstatt /tmp lauten muss.

Version 4.4.0

Veröffentlichungsdatum: 14. März 2016

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- 1.0.0 wurde hinzugefügt HCatalog
- Sqoop-Sandbox 1.4.6 wurde hinzugefügt.
- Upgrade auf Presto 0.136
- Upgrade auf Zeppelin 0.5.6
- Upgrade auf Mahout 0.11.1
- `dynamicResourceAllocation` wurde standardmäßig aktiviert.
- Es wurde eine Tabelle mit allen Konfigurationsklassifizierungen für die Version hinzugefügt. Weitere Informationen finden Sie in der Konfigurationsklassifizierungstabelle unter [Konfigurieren von Anwendungen](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die `maximizeResourceAllocation` Einstellung nicht genug Speicher für YARN ApplicationMaster Daemons reservierte.

- Es wurde ein Problem behoben, das bei einem benutzerdefinierten Gerät aufgetreten ist. DNS Wenn Einträge in `resolve.conf` den angegebenen, benutzerdefinierten Einträge vorangestellt werden, können die benutzerdefinierten Einträge nicht aufgelöst werden. Dieses Verhalten wurde durch Cluster in einem beeinflusst, in VPC dem der VPC Standard-Nameserver als oberster Eintrag eingefügt wurde `resolve.conf`.
- Es wurde ein Problem behoben, bei dem die Python-Standardversion auf Version 2.7 geändert wurde und boto für diese Version nicht installiert war.
- Es wurde ein Problem behoben, bei dem YARN Container und Spark-Anwendungen eine eindeutige Ganglia-Round-Robin-Datenbankdatei (RRD) generierten, was dazu führte, dass die erste Festplatte, die an die Instanz angeschlossen war, voll wurde. Aufgrund dieses Fixes wurden Metriken auf YARN Containerebene und Kennzahlen auf Spark-Anwendungsebene deaktiviert.
- Es wurde ein Problem im Protokoll-Pusher behoben, bei dem alle leeren Protokollordner gelöscht wurden. Das hatte zur Folge, dass Hive CLI nicht protokollieren konnte, weil der Log-Pusher den leeren `user` Ordner darunter entfernte. `/var/log/hive`
- Es wurde ein Problem mit Hive-Importen behoben, da sich auf die Partitionierung auswirkte und beim Import zu einem Fehler führte.
- Es wurde ein Problem behoben, bei dem EMRFS s3-dist-cp Bucket-Namen, die Punkte enthielten, nicht richtig behandelte.
- Das Verhalten wurde EMRFS dahingehend geändert, dass in Buckets mit aktivierter Versionierung die `_${folder}` Markierungsdatei nicht kontinuierlich erstellt wird, was zu einer verbesserten Leistung für Buckets mit Versionierung beitragen kann.
- Das Verhalten wurde EMRFS dahingehend geändert, dass es keine Befehlsdateien verwendet, außer in Fällen, in denen die clientseitige Verschlüsselung aktiviert ist. Wenn Sie Anweisungsdateien bei Verwendung der clientseitigen Verschlüsselung löschen möchten, können Sie die `emrfs-site.xml`-Eigenschaft, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, auf "true" festlegen.
- Die YARN Protokollaggregation wurde dahingehend geändert, dass die Protokolle zwei Tage lang am Aggregationsziel aufbewahrt werden. Das Standardziel ist der Speicher Ihres ClustersHDFS. Wenn Sie diese Dauer ändern möchten, ändern Sie den Wert `yarn.log-aggregation.retain-seconds` mit der Konfigurationsklassifizierung `yarn-site` beim Erstellen Ihres Clusters. Wie immer können Sie Ihre Anwendungsprotokolle beim Erstellen Ihres Clusters mit dem Parameter `log-uri` in Amazon S3 speichern.

Angewendete Patches

Die folgenden Patches aus Open-Source-Projekten sind in dieser Version enthalten:

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Version 4.3.0

Veröffentlichungsdatum: 19. Januar 2016

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Upgrade auf Hadoop 2.7.1
- Upgrade auf Spark 1.6.0
- Upgrade auf Ganglia 3.7.2
- Upgrade auf Presto 0.130

Amazon EMR hat einige Änderungen daran vorgenommen, `spark.dynamicAllocation.enabled` wenn es auf `true` gesetzt ist; es ist standardmäßig falsch. Wenn die Einstellung auf `"true"` festgelegt ist, wirkt sich dies auf von der Einstellung `maximizeResourceAllocation` vorgenommene Standardeinstellungen aus:

- Wenn `spark.dynamicAllocation.enabled` auf `"true"` eingestellt ist, wird `spark.executor.instances` nicht von `maximizeResourceAllocation` festgelegt.
- Die Einstellung `spark.driver.memory` wird nun basierend auf den Instance-Typen im Cluster auf ähnliche Weise konfiguriert, wie `spark.executors.memory` festgelegt wird. Da die Spark-Treiberanwendung jedoch entweder auf der Master- oder auf einer der Core-Instances ausgeführt werden kann (z. B. im YARN Client- bzw. Cluster-Modus), wird die `spark.driver.memory` Einstellung auf der Grundlage des Instance-Typs des kleineren Instance-Typs zwischen diesen beiden Instance-Gruppen festgelegt.
- Die `spark.default.parallelism` Einstellung ist jetzt auf die doppelte Anzahl der für YARN Container verfügbaren CPU Kerne festgelegt. In früheren Versionen war dies die Hälfte des Werts.

- Die Berechnungen für den für YARN Spark-Prozesse reservierten Speicheraufwand wurden angepasst, sodass sie genauer sind, was zu einer geringfügigen Erhöhung der Gesamtspeichermenge führte, die Spark zur Verfügung steht (d. h. `spark.executor.memory`).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- YARN Die Protokollaggregation ist jetzt standardmäßig aktiviert.
- Es wurde ein Problem behoben, bei dem Protokolle nicht in den Amazon S3 S3-Log-Bucket eines Clusters übertragen wurden, wenn die YARN Protokollaggregation aktiviert war.
- YARN Für Containergrößen gilt jetzt ein neuer Mindestwert von 32 für alle Knotentypen.
- Es wurde ein Problem mit Ganglia behoben, das zu übermäßigen Festplatten-E/A-Vorgängen auf dem Master-Knoten in großen Clustern führte.
- Es wurde ein Problem behoben, das verhinderte, dass Anwendungsprotokolle per Push-Verfahren an Amazon S3 übertragen wurden, wenn ein Cluster heruntergefahren wird.
- Es wurde ein Problem behoben EMRFSCLI, das dazu führte, dass bestimmte Befehle fehlschlügen.
- Es wurde ein Problem mit Zeppelin behoben, das verhinderte, dass Abhängigkeiten in die Basisdatei geladen wurden. `SparkContext`
- Es wurde ein Problem behoben, das aus einer Größenanpassung resultierte bei dem Versuch, Instances hinzuzufügen.
- Es wurde ein Problem in Hive behoben, bei dem `CREATE TABLE AS` übermäßig `SELECT` viele Listenaufrufe an Amazon S3 tätigte.
- Es wurde ein Problem behoben, bei dem große Cluster nicht ordnungsgemäß bereitgestellt wurden wenn Hue, Oozie und Ganglia installiert sind.
- Es wurde ein Problem in `s3-dist-cp` behoben, bei dem ein Beendigungscode gleich Null zurückgegeben wurde, wenn ein Fehler auftrat.

Angewendete Patches

Die folgenden Patches aus Open-Source-Projekten sind in dieser Version enthalten:

- [OOZIE-2402](#)
- [HIVE-12502](#)
- [HIVE-10631](#)

- [HIVE-12213](#)
- [HIVE-10559](#)
- [HIVE-12715](#)
- [HIVE-10685](#)

Version 4.2.0

Veröffentlichungsdatum: 18. November 2015

Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Ganglia-Support wurde hinzugefügt.
- Upgrade auf Spark 1.5.2
- Upgrade auf Presto 0.125
- Upgrade auf Oozie 4.2.0
- Upgrade auf Zeppelin 0.5.5
- Auf 1.10.27 AWS SDK for Java aktualisiert

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben EMRFSCLI, bei dem der Standardname der Metadatentabelle nicht verwendet wurde.
- Es wurde ein Problem behoben, das bei der Verwendung von Tabellen mit ORC -back in Amazon S3 auftrat.
- Es wurde ein Problem behoben, das bei einem Python-Versionskonflikt in der Spark-Konfiguration auftrat.
- Es wurde ein Problem behoben, bei dem der Status eines YARN Knotens aufgrund von DNS Problemen mit Clustern in einem VPC nicht gemeldet wurde.
- Es wurde ein Problem behoben, das YARN auftrat, wenn Knoten außer Betrieb genommen wurden, was dazu führte, dass Anwendungen hängenblieben oder neue Anwendungen nicht geplant werden konnten.
- Es wurde ein Problem behoben, das auftrat, wenn Cluster mit dem Status TIMED __ OUT beendet wurden. STARTING

- Es wurde ein Problem behoben, das auftrat, wenn die EMRFS Scala-Abhängigkeit in andere Builds aufgenommen wurde. Die Scala-Abhängigkeit wurde entfernt.

Anwendungen konfigurieren

Um die Standardkonfigurationen für eine Anwendung zu überschreiben, können Sie ein Konfigurationsobjekt bereitstellen. Sie können entweder eine Kurzsyntax verwenden, um die Konfiguration bereitzustellen, oder Sie können das Konfigurationsobjekt in einer Datei referenzieren. JSON Konfigurationsobjekte bestehen aus einer Klassifizierung, Eigenschaften und optionalen verschachtelten Konfigurationen. Die Eigenschaften entsprechen den Anwendungseinstellungen, die Sie ändern möchten. Sie können mehrere Klassifizierungen für mehrere Anwendungen in einem einzigen Objekt angeben. JSON

Warning

Die Amazon EMR Describe- und API List-Operationen geben benutzerdefinierte und konfigurierbare Einstellungen, die als Teil der EMR Amazon-Auftragsabläufe verwendet werden, im Klartext aus. Informationen zur Angabe vertraulicher Informationen wie Kennwörter in diesen Einstellungen finden Sie unter [Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern](#).

Die verfügbaren Konfigurationsklassifizierungen variieren je nach EMR Amazon-Release-Version. Eine Liste der Konfigurationsklassifizierungen, die in einer bestimmten Release-Version unterstützt werden, finden Sie auf der Seite für diese Release-Version unter [Über Amazon EMR Releases](#).

Im Folgenden finden Sie eine JSON Beispieldatei für eine Liste von Konfigurationen.

```
[
  {
    "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"
    }
  }
]
```



```
}
]
```

Eine Konfigurationsklassifizierung bezieht sich häufig auf eine anwendungsspezifische Konfigurationsdatei. Die `hive-site`-Klassifizierung ist Einstellungen in der `hive-site.xml`-Konfigurationsdatei für Hive zugeordnet. Eine Ausnahme bildet die veraltete Bootstrap-Aktion `configure-daemons`, mit der Umgebungsparameter wie z. B. `--namenode-heap-size` definiert werden. Optionen wie diese sind in den Klassifizierungen `hadoop-env` und `yarn-env` mit ihren eigenen verschachtelten Exportklassifizierungen zusammengefasst. Verwenden Sie die Export-Unterklassifizierung, wenn eine Klassifizierung auf `env` endet.

Eine weitere Ausnahme ist `s3get`. Sie wurde verwendet, um ein `EncryptionMaterialsProvider`-Kundenobjekt auf jedem Knoten in einem Cluster für die Verwendung bei der clientseitigen Verschlüsselung zu platzieren. Der Klassifizierung `emrfs-site` wurde zu diesem Zweck eine Option hinzugefügt.

Im Folgenden finden Sie ein Beispiel für die `hadoop-env`-Klassifizierung.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Im Folgenden finden Sie ein Beispiel für die `yarn-env`-Klassifizierung.

```
[
  {
    "Classification": "yarn-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdp:transport=dt_socket"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Die folgenden Einstellungen gehören nicht zu einer Konfigurationsdatei, sondern werden von Amazon verwendet EMR, um möglicherweise mehrere Einstellungen in Ihrem Namen zu konfigurieren.

Von Amazon kuratierte Einstellungen EMR

Anwendung	Klassifizierung der Versionsbezeichnung	Gültige Eigenschaften	Wann sollte dies verwendet werden?
Spark	spark	maximizeResourceAllocation	Konfigurieren von Executors zur Nutzung der maximalen Ressourcen eines jeden Knotens

Themen

- [Anwendungen bei der Clustererstellung konfigurieren](#)
- [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#)
- [Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern](#)

- [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#)

Anwendungen bei der Clustererstellung konfigurieren

Wenn Sie einen Cluster erstellen, können Sie die Standardkonfigurationen für Anwendungen mithilfe der EMR Amazon-Konsole, der AWS Command Line Interface (AWS CLI) oder der überschreiben AWS SDK.

Um die Standardkonfiguration für eine Anwendung zu überschreiben, geben Sie benutzerdefinierte Werte in einer Konfigurationsklassifizierung an. Eine Konfigurationsklassifizierung entspricht einer XML Konfigurationsdatei für eine Anwendung, wie `hive-site.xml` z.

Die Konfigurationsklassifizierungen variieren je nach EMR Amazon-Release-Version. Eine Liste der Konfigurationsklassifizierungen, die in einer bestimmten Release-Version verfügbar sind, finden Sie auf der Release-Detailseite. Zum Beispiel [EMRAmazon-Version 6.4.0](#).

Geben Sie beim Erstellen eines Clusters eine Konfiguration in der Konsole an

Um eine Konfiguration bereitzustellen, navigieren Sie zur Seite Cluster erstellen und erweitern Sie Softwareeinstellungen. Sie können die Konfiguration dann direkt eingeben, indem Sie entweder JSON oder eine Kurzsyntax verwenden, die im Shadow-Text in der Konsole veranschaulicht wird. Andernfalls können Sie Amazon S3 URI für eine Datei mit einem JSON Configurations Objekt bereitstellen.

Um eine Konfiguration für eine Instance-Gruppe bereitzustellen, wählen Sie einen Cluster in Ihrer Clusterliste aus und wählen Sie dann den Tab Konfigurationen. Wählen Sie in der Tabelle mit den Instanzgruppenkonfigurationen die zu bearbeitende Instanzgruppe aus und klicken Sie dann auf Neu konfigurieren.

Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie

Sie können eine Konfiguration für bereitstellen, `create-cluster` indem Sie einen Pfad zu einer lokal oder in Amazon S3 gespeicherten JSON Datei angeben. Im folgenden Beispiel wird davon ausgegangen, dass Sie Standardrollen für Amazon verwenden EMR und dass die Rollen erstellt wurden. Wenn Sie die Rollen erstellen müssen, führen Sie zuerst `aws emr create-default-roles` aus.

Wenn sich die Konfiguration im lokalen Verzeichnis befindet, können Sie den folgenden Beispielbefehl verwenden.

```
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
```

Wenn sich Ihre Konfiguration in einem Amazon-S3-Pfad befindet, müssen Sie die folgende Problemumgehung einrichten, bevor Sie den Amazon-S3-Pfad an den `create-cluster`-Befehl übergeben.

```
#!/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
```

Geben Sie eine Konfiguration mit Java an SDK, wenn Sie einen Cluster erstellen

Der folgende Programmausschnitt zeigt, wie eine Konfiguration mit AWS SDK for Java bereitgestellt wird.

```
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")
    );
```

Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Anwendungen neu konfigurieren und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Dazu können Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK.

Wenn Sie eine Anwendungskonfiguration für eine Instance-Gruppe in der neuen EMR Amazon-Konsole aktualisieren, versucht die Konsole, die neue Konfiguration mit der vorhandenen Konfiguration zusammenzuführen, um eine neue, aktive Konfiguration zu erstellen. In dem ungewöhnlichen Fall, dass Amazon die Konfiguration nicht zusammenführen EMR kann, warnt Sie die Konsole.

Nachdem Sie eine Rekonfigurationsanfrage für eine Instance-Gruppe eingereicht haben, EMR weist Amazon der neuen Konfigurationsspezifikation eine Versionsnummer zu. Sie können die Versionsnummer einer Konfiguration oder den Status einer Instanzgruppe verfolgen, indem Sie sich die CloudWatch Ereignisse ansehen. Weitere Informationen finden Sie unter [CloudWatch Ereignisse überwachen](#).

Note

Sie können Clusterkonfigurationen, die bei der Clustererstellung angegeben wurden, nur überschreiben und nicht löschen. Wenn es Unterschiede zwischen der vorhandenen Konfiguration und der von Ihnen bereitgestellten Datei gibt, EMR setzt Amazon manuell

geänderte Konfigurationen, z. B. Konfigurationen, die Sie geändert haben, während Sie mit Ihrem Cluster verbunden waren SSH, auf die Cluster-Standard-Einstellungen für die angegebene Instance-Gruppe zurück.

Überlegungen bei der Neukonfiguration einer Instance-Gruppe

Aktionen zur Neukonfiguration

Wenn Sie eine Rekonfigurationsanfrage über die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die einreichen AWS SDK, EMR überprüft Amazon die vorhandene Cluster-Konfigurationsdatei. Wenn es Unterschiede zwischen der vorhandenen Konfiguration und der von Ihnen bereitgestellten Datei gibt, EMR leitet Amazon Rekonfigurationsaktionen ein, startet einige Anwendungen neu und setzt alle manuell geänderten Konfigurationen, wie Konfigurationen, die Sie geändert haben, während Sie mit Ihrem Cluster verbunden waren SSH, auf die Cluster-Standard-Einstellungen für die angegebene Instance-Gruppe zurück.

Note

Amazon EMR führt bei jeder Neukonfiguration der Instanzgruppe einige Standardaktionen durch. Diese Standardaktionen können zu Konflikten mit den von Ihnen vorgenommenen Cluster-Anpassungen führen und zu Fehlern bei der Neukonfiguration führen. Weitere Informationen, wie Sie bei fehlgeschlagener Neukonfiguration beheben können, finden Sie unter [Probleme bei der Neukonfiguration von Instance-Gruppen beheben](#).

Amazon leitet EMR auch Neukonfigurationsaktionen für die Konfigurationsklassifizierungen ein, die Sie in Ihrer Anfrage angeben. Eine vollständige Liste dieser Aktionen finden Sie im Abschnitt Konfigurationsklassifizierungen für die Version von Amazon EMR, die Sie verwenden. Zum Beispiel [6.2.0.-Konfigurationsklassifizierungen](#).

Note

Der EMR Amazon-Versionshandbuch listet nur Neukonfigurationsaktionen ab den EMR Amazon-Versionen 5.32.0 und 6.2.0 auf.

Betriebsunterbrechung

Amazon EMR folgt einem fortlaufenden Prozess zur Neukonfiguration von Instances in den Instance-Gruppen Task und Core. Zu einem bestimmten Zeitpunkt werden nur 10 % der Instances in einer Instance-Gruppe geändert und neu gestartet. So benötigt der Prozess mehr Zeit, gleichzeitig ist aber die Wahrscheinlichkeit eines Anwendungsausfalls in einem aktiven Cluster geringer.

Um YARN Jobs während eines YARN Neustarts auszuführen, können Sie entweder einen EMR Amazon-Cluster mit mehreren Master-Knoten erstellen oder `true` in Ihrer `yarn-site` Konfigurationsklassifizierung `yarn.resourcemanager.recovery.enabled` auf einstellen. Weitere Informationen zur Verwendung mehrerer Master-Knoten finden Sie unter [Hochverfügbarkeit YARN ResourceManager](#).

Validierung der Anwendung

Amazon EMR überprüft nach dem Neustart der Neukonfiguration, ob jede Anwendung auf dem Cluster ausgeführt wird. Wenn eine Anwendung nicht verfügbar ist, schlägt der gesamte Neukonfigurationsvorgang fehl. Wenn ein Neukonfigurationsvorgang fehlschlägt, EMR macht Amazon die Konfigurationsparameter auf die vorherige funktionierende Version rückgängig.

Note

Um Fehler bei der Neukonfiguration zu vermeiden, empfehlen wir, nur die Anwendungen auf Ihrem Cluster zu installieren, die Sie verwenden möchten. Wir empfehlen Ihnen außerdem, sicherzustellen, dass alle Clusteranwendungen fehlerfrei sind und ausgeführt werden, bevor Sie eine Rekonfigurationsanfrage einreichen.

Arten der Neukonfiguration

Sie können eine Instance-Gruppe auf zwei Arten neu konfigurieren:

- **Überschreiben.** Standardmethode für die Rekonfiguration und die einzige, die in EMR Amazon-Versionen vor 5.35.0 und 6.6.0 verfügbar ist. Diese Rekonfigurationsmethode überschreibt wahllos alle Dateien im Cluster mit dem neu eingereichten Konfigurationssatz. Die Methode löscht alle Änderungen an den Konfigurationsdateien, die außerhalb der Rekonfiguration vorgenommen wurden. API
- **Mischen.** Die Rekonfigurationsmethode wird für die EMR Amazon-Versionen 5.35.0 und 6.6.0 und höher unterstützt, außer von der EMR Amazon-Konsole aus, wo sie von keiner Version

unterstützt wird. Diese Rekonfigurationsmethode führt die neu eingereichten Konfigurationen mit Konfigurationen zusammen, die bereits auf dem Cluster vorhanden sind. Diese Option fügt nur die neuen Konfigurationen hinzu oder ändert sie, die Sie einreichen. Sie behält bestehende Konfigurationen bei.

 Note

Amazon überschreibt EMR weiterhin einige wichtige Hadoop-Konfigurationen, die es benötigt, um sicherzustellen, dass der Service ordnungsgemäß ausgeführt wird.

Einschränkungen

Beachten Sie bei der Neukonfiguration einer Instance-Gruppe in einem laufenden Cluster die folgenden Einschränkungen:

- YARN-Anwendungen können beim Neustart fehlschlagen oder Clusterprobleme verursachen, insbesondere wenn die Anwendungen nicht richtig konfiguriert sind. Bei Clustern, die sich dem maximalen Arbeitsspeicher und der maximalen CPU Auslastung nähern, können nach dem Neustart Probleme auftreten. Dies gilt insbesondere für die Master-Instance-Gruppe.
- Sie können keine Rekonfigurationsanfrage stellen, wenn die Größe einer Instance-Gruppe geändert wird. Wenn eine Rekonfiguration initiiert wird, während die Größe einer Instance-Gruppe geändert wird, kann die Rekonfiguration erst durchgeführt werden, nachdem die Größenänderung abgeschlossen wurde (und umgekehrt).
- Nach der Neukonfiguration einer Instanzgruppe EMR startet Amazon die Anwendungen neu, damit die neuen Konfigurationen wirksam werden können. Auftragsfehler und anderes unerwartetes Anwendungsverhalten kann auftreten, wenn die Anwendungen während der Rekonfiguration genutzt werden.
- Wenn eine Neukonfiguration für eine Instanzgruppe fehlschlägt, EMR macht Amazon die Konfigurationsparameter auf die vorherige funktionierende Version rückgängig. Schlägt auch diese Wiederherstellung fehl, müssen Sie eine neue `ModifyInstanceGroup`-Anforderung zur Wiederherstellung der Instance-Gruppe aus dem Status `SUSPENDED` senden.
- Rekonfigurationsanfragen für Phoenix-Konfigurationsklassifizierungen werden nur in EMR Amazon-Version 5.23.0 und höher unterstützt und werden in EMR Amazon-Version 5.21.0 oder 5.22.0 nicht unterstützt.

- Rekonfigurationsanfragen für HBase Konfigurationsklassifizierungen werden nur in EMR Amazon-Version 5.30.0 und höher unterstützt und werden in den EMR Amazon-Versionen 5.23.0 bis 5.29.0 nicht unterstützt.
- Amazon EMR unterstützt Anfragen zur Neukonfiguration von Anwendungen auf einem EMR Amazon-Cluster mit mehreren Primärknoten nur in EMR Amazon-Versionen 5.27.0 und höher.
- Die Neukonfiguration der `hdfs-encryption-zones` Klassifizierung oder einer der KMS Hadoop-Konfigurationsklassifizierungen wird auf einem EMR Amazon-Cluster mit mehreren Primärknoten nicht unterstützt.
- Amazon unterstützt EMR derzeit bestimmte Rekonfigurationsanfragen für den Capacity Scheduler nicht, die einen Neustart des erfordern. YARN ResourceManager Beispielsweise können Sie eine Warteschlange nicht vollständig entfernen.

Eine Instance-Gruppe in der Konsole neu konfigurieren

Note

Die EMR Amazon-Konsole unterstützt keine Rekonfigurationen vom Typ Merge.

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>
2. Wählen Sie in der Cluster-Liste den aktiven Cluster, der umkonfiguriert werden soll, unter Name.
3. Öffnen Sie die Seite mit den Cluster-Details für den Cluster und wechseln Sie zur Registerkarte Konfigurationen.
4. Wählen Sie in der Dropdown-Liste Filter die Instance-Gruppe aus, die rekonfiguriert werden soll.
5. Wählen Sie im Drop-down-Menü „Neu konfigurieren“ entweder „In Tabelle bearbeiten“ oder „In JSON Datei bearbeiten“.
 - In Tabelle bearbeiten: Bearbeiten Sie in der Tabelle der Konfigurationsklassifizierungen Eigenschaft und Wert vorhandener Konfigurationen oder wählen Sie Konfiguration hinzufügen, um für weitere Konfigurationsklassifizierungen anzugeben.
 - In JSON Datei bearbeiten — Geben Sie die Konfiguration direkt einJSON, oder verwenden Sie eine Kurzsyntax (demonstriert im Schattentext). Andernfalls stellen Sie Amazon S3 URI für eine Datei mit einem JSON Configurations Objekt bereit.

Note

Die Spalte Quelle in der Tabelle der Konfigurationsklassifizierungen zeigt, ob die Konfiguration beim Erstellen eines Clusters oder beim Angeben zusätzlicher Konfigurationen für diese Instance-Gruppe bereitgestellt wird. Sie können Konfigurationen für eine Instance-Gruppe aus beiden Quellen bearbeiten. Sie können die ursprünglichen Cluster-Konfigurationen nicht löschen, aber sie für eine Instance-Gruppe überschreiben.

Sie können außerdem verschachtelte Konfigurationsklassifizierungen direkt in der Tabelle hinzufügen oder bearbeiten. Um beispielsweise eine zusätzliche `export-` Unterklassifizierung von `hadoop-env` bereitzustellen, fügen Sie eine `hadoop.export-` Konfigurationsklassifizierung in die Tabelle ein. Anschließend stellen Sie eine Eigenschaft samt Wert für diese Klassifizierung bereit.

6. (Optional) Wählen Sie `Apply this configuration to all active instance groups` (Diese Konfiguration allen aktiven Instance-Gruppen zuweisen).
7. Speichern Sie die Änderungen.

Rekonfigurieren Sie eine Instanzgruppe mit dem CLI

Verwenden Sie den Befehl `modify-instance-groups`, um eine neue Konfiguration für eine Instance-Gruppe in einem laufenden Cluster anzugeben.

Note

Ersetzen Sie in den folgenden Beispielen `<j-2AL4XXXXXX5T9>` durch Ihre Cluster-ID und ersetzen `<ig-1xxxxxxx9>` mit Ihrer Instanzgruppen-ID.

Example – Eine Konfiguration für eine Instance-Gruppe ersetzen

Das folgende Beispiel verweist auf eine JSON Konfigurationsdatei, die aufgerufen wird `instanceGroups.json`, um die Eigenschaft der YARN NodeManager Festplattenintegritätsprüfung für eine Instanzgruppe zu bearbeiten.

1. Bereiten Sie Ihre Konfigurationsklassifizierung vor und speichern Sie sie als `instanceGroups.json` in demselben Verzeichnis, in dem Sie den Befehl ausführen werden.

```
[
  {
    "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. Führen Sie den folgenden Befehl aus.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Einer Instance-Gruppe eine Konfiguration hinzufügen

Wenn Sie einer Instance-Gruppe eine Konfiguration hinzufügen möchten, müssen Sie alle zuvor angegebenen Konfigurationen für diese Instance-Gruppe in Ihre neue `ModifyInstanceGroup`-Anforderung aufnehmen. Andernfalls werden die zuvor für die angegebenen Konfigurationen entfernt.

Im folgenden Beispiel wird eine Eigenschaft für den YARN NodeManager Virtual Memory Checker hinzugefügt. Die Konfiguration enthält auch zuvor angegebene Werte für die YARN NodeManager Festplattenintegritätsprüfung, sodass die Werte nicht überschrieben werden.

1. Bereiten Sie den folgenden Inhalt in `instanceGroups.json` vor und speichern Sie ihn im selben Verzeichnis, in dem Sie den Befehl ausführen werden.

```
[
  {
    "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. Führen Sie den folgenden Befehl aus.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Einer Instance-Gruppe mit Rekonfiguration vom Typ Zusammenführen eine Konfiguration hinzufügen

Wenn Sie die Standardmethode Rekonfiguration Überschreiben verwenden möchten, um eine Konfiguration hinzuzufügen, müssen Sie alle zuvor angegebenen Konfigurationen für diese Instance-Gruppe in Ihre neue ModifyInstanceGroup-Anfrage aufnehmen. Andernfalls entfernt Überschreiben die Konfigurationen, die Sie zuvor angegeben haben. Bei der Neukonfiguration mit Zusammenführen müssen Sie dies nicht tun. Stattdessen müssen Sie sicherstellen, dass Ihre Anfrage nur die neuen Konfigurationen enthält.

Im folgenden Beispiel wird eine Eigenschaft für den YARN NodeManager Virtual Memory Checker hinzugefügt. Da es sich um eine Rekonfiguration vom Typ Merge handelt, werden zuvor angegebene Werte für die YARN NodeManager Festplattenintegritätsprüfung nicht überschrieben.

1. Bereiten Sie den folgenden Inhalt in `instanceGroups.json` vor und speichern Sie ihn im selben Verzeichnis, in dem Sie den Befehl ausführen werden.

```
[
  {"InstanceGroupId": "<ig-1xxxxxxx9>",
   "ReconfigurationType": "MERGE",
   "Configurations": [
     {"Classification": "yarn-site",
      "Properties": {
        "yarn.nodemanager.vmem-check-enabled": "true",
        "yarn.nodemanager.vmem-pmem-ratio": "3.0"
      }
    },
     "Configurations": []
   ]
 }
 ]
```

2. Führen Sie den folgenden Befehl aus.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Eine Konfiguration für eine Instance-Gruppe löschen

Um eine Konfiguration für eine Instance-Gruppe zu löschen, senden Sie eine neue Neukonfigurationsanforderung, die die vorherige Konfiguration ausschließt.

Note

Sie können die ursprüngliche Cluster-Konfiguration nur überschreiben. Sie kann nicht gelöscht werden.

Um beispielsweise die Konfiguration für die YARN NodeManager Festplattenintegritätsprüfung aus dem vorherigen Beispiel zu löschen, reichen Sie eine neue `instanceGroups.json` mit dem folgenden Inhalt ein.

```
[
  {
```

```

    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]

```

Note

Um alle Konfigurationen der letzten Rekonfigurationsanforderung zu löschen, senden Sie eine Rekonfigurationsanforderung mit einem leeren Konfigurations-Array. Zum Beispiel

```

[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": []
  }
]

```

Example – Eine Instance-Gruppe in einer Anfrage neu konfigurieren oder die Größe ändern

Das folgende Beispiel JSON zeigt, wie Sie eine Instanzgruppe in derselben Anfrage neu konfigurieren und ihre Größe ändern können.

```

[
  {
    "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"
      }
    }
  ]
}
]

```

Rekonfigurieren Sie eine Instanzgruppe mithilfe von Java SDK

Note

Ersetzen Sie in den folgenden Beispielen `<j-2AL4XXXXXX5T9>` durch Ihre Cluster-ID und ersetzen `<ig-1xxxxxxx9>` mit Ihrer Instanzgruppen-ID.

Der folgende Codeausschnitt bietet eine neue Konfiguration für eine Instance-Gruppe unter Verwendung von 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);

```

Der folgende Codeausschnitt löscht eine zuvor angegebene Konfiguration für eine Instance-Gruppe, indem ein leeres Array von Konfigurationen bereitgestellt wird.

```

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);

```

Probleme bei der Neukonfiguration von Instance-Gruppen beheben

Wenn der Neukonfigurationsprozess für eine Instanzgruppe fehlschlägt, macht Amazon die Neukonfiguration EMR rückgängig und protokolliert eine Fehlermeldung mithilfe eines Amazon-Ereignisses. CloudWatch Das Ereignis bietet eine kurze Zusammenfassung des Neukonfigurationsfehlers. Es listet die Instances auf, bei denen die Rekonfiguration fehlgeschlagen ist, sowie die entsprechenden Fehlermeldungen. Das Folgende ist ein Beispiel für eine Fehlermeldung.

```

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".
```

Weitere Daten zu einem Fehler bei der Rekonfiguration können Sie in den Protokollen zur Knotenbereitstellung nachlesen. Dies ist besonders nützlich, wenn Sie eine Nachricht wie die folgende erhalten.

```
i-xxxxxxx1 failed with message "Unable to complete transaction and some changes were
applied."
```

On the node

So greifen Sie auf die Bereitstellungsprotokolle von Knoten zu, indem Sie eine Verbindung zu einem Knoten herstellen

1. Wird verwendet SSH, um eine Verbindung zu dem Knoten herzustellen, auf dem die Neukonfiguration fehlgeschlagen ist. Anweisungen finden Sie unter [Connect zu Ihrer Linux-Instance](#) herstellen im EC2 Amazon-Benutzerhandbuch für Linux-Instances.
2. Navigieren Sie zum folgenden Verzeichnis, das die Protokolldateien für die Knotenbereitstellung enthält.

```
/mnt/var/log/provision-node/
```

3. Öffnen Sie das `reports`-Unterverzeichnis und suchen Sie nach dem Knotenbereitstellungsbericht für Ihre Neukonfiguration. Das `reports` Verzeichnis organisiert Protokolle nach Versionsnummer der Rekonfiguration, Universally Unique Identifier (UUID), EC2 Amazon-Instance-IP-Adresse und Zeitstempel. Jeder Bericht ist eine komprimierte YAML Datei, die detaillierte Informationen über den Rekonfigurationsprozess enthält.

Im Folgenden finden Sie ein Beispiel für einen Berichts-Dateiname und einen Pfad.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxx/ip-10-73-xxx-
xxx.ec2.internal/202104061715.yaml.gz
```

4. Sie können einen Bericht mit einem Dateibetrachter wie `zless` im folgenden Beispiel untersuchen.

```
zless 202104061715.yaml.gz
```

Amazon S3

So greifen Sie mit Amazon S3 auf Knotenbereitstellungsprotokolle zu

1. Melden Sie sich bei der an AWS Management Console und öffnen Sie die Amazon S3 S3-Konsole unter <https://console.aws.amazon.com/s3/>.
2. Öffnen Sie den Amazon-S3-Bucket, den Sie angegeben haben, als Sie den Cluster für die Archivierung von Protokolldateien in konfiguriert haben.
3. Navigieren Sie zum folgenden Ordner, der die Protokolldateien für die Knotenbereitstellung enthält:

```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-  
node/
```

4. Öffnen Sie den `reports`-Ordner und suchen Sie nach dem Bericht zur Knotenbereitstellung für Ihre Rekonfiguration. Der `reports` Ordner organisiert die Protokolle nach Versionsnummer der Rekonfiguration, Universally Unique Identifier (UUID), EC2 Amazon-Instance-IP-Adresse und Zeitstempel. Jeder Bericht ist eine komprimierte YAML Datei, die detaillierte Informationen über den Rekonfigurationsprozess enthält.

Im Folgenden finden Sie ein Beispiel für einen Berichts-Dateiname und einen Pfad.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxxx/ip-10-73-xxx-  
xxx.ec2.internal/202104061715.yaml.gz
```

5. Um eine Protokolldatei anzuzeigen, können Sie sie als Textdatei von Amazon S3 auf Ihren lokalen Computer herunterladen. Anweisungen finden Sie unter [Objekt herunterladen](#).

Jede Protokolldatei enthält einen detaillierten Bereitstellungsbericht für die zugehörige Rekonfiguration. Um Informationen zu Fehlermeldungen zu finden, können Sie nach der `err`-Protokollebene eines Berichts suchen. Das Berichtsformat hängt von der Version von Amazon in EMR Ihrem Cluster ab.

Das folgende Beispiel zeigt Fehlerinformationen für EMR Amazon-Release-Versionen vor 5.32.0 und 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
```

Die EMR Amazon-Versionen 5.32.0 und 6.2.0 und höher verwenden stattdessen das folgende Format.

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
    - err
  time: '2021-01-01 00:00:00.000000 +00:00'
  file:
  line:
```

Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern

Die Amazon EMR Describe- und API List-Operationen, die benutzerdefinierte Konfigurationsdaten (wie DescribeCluster und ListInstanceGroups) ausgeben, tun dies im Klartext. Amazon EMR integriert sich mit AWS Secrets Manager sodass Sie Ihre Daten in Secrets Manager speichern und das Geheimnis ARN in Ihren Konfigurationen verwenden können. Auf diese Weise geben Sie vertrauliche Konfigurationsdaten nicht EMR im Klartext an Amazon weiter und stellen sie externen APIs Benutzern zur Verfügung. Wenn Sie angeben, dass ein Schlüssel-Wert-Paar ein in Secrets Manager gespeichertes Geheimnis ARN für ein Geheimnis enthält, EMR ruft Amazon dieses Geheimnis ab, wenn es Konfigurationsdaten an den Cluster sendet. Amazon sendet die Anmerkung EMR nicht, wenn es External APIs zur Anzeige der Konfiguration verwendet.

Ein Secret erstellen

Um ein Geheimnis zu erstellen, befolgen Sie die Schritte unter [AWS Secrets Manager -Geheimnis erstellen](#) im AWS Secrets Manager -Benutzerhandbuch. In Schritt 3 müssen Sie das Klartext-Feld auswählen, um Ihren sensiblen Wert einzugeben.

Beachten Sie, dass Secrets Manager zwar zulässt, dass ein Geheimnis bis zu 65536 Byte enthält, Amazon jedoch die kombinierte Länge des Eigenschaftsschlüssels (ohne Anmerkung) und des abgerufenen geheimen Werts auf 1024 Zeichen EMR begrenzt.

Gewähren EMR Sie Amazon Zugriff, um das Geheimnis abzurufen

Amazon EMR verwendet eine IAM Service-Rolle, um Cluster für Sie bereitzustellen und zu verwalten. Die Servicerolle für Amazon EMR definiert die zulässigen Aktionen für Amazon, EMR wenn Amazon Ressourcen bereitstellt und Service-Level-Aufgaben ausführt, die nicht im Kontext einer EC2 Amazon-Instance ausgeführt werden, die innerhalb eines Clusters ausgeführt wird. Weitere Informationen zu Servicerollen finden Sie unter [Servicerolle für Amazon EMR \(EMRRolle\)](#) und [IAMRollen anpassen](#).

Damit Amazon EMR den geheimen Wert aus Secrets Manager abrufen kann, fügen Sie Ihrer EMR Amazon-Rolle die folgende Richtlinienerklärung hinzu, wenn Sie Ihren Cluster starten.

```
{
  "Sid": "AllowSecretsRetrieval",
  "Effect": "Allow",
  "Action": "secretsmanager:GetSecretValue",
  "Resource": [
    "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  ]
}
```

Wenn Sie das Geheimnis mit einem vom Kunden verwalteten Schlüssel erstellen AWS KMS key, müssen Sie der EMR Amazon-Rolle auch eine `kms:Decrypt` Berechtigung für den von Ihnen verwendeten Schlüssel hinzufügen. Weitere Informationen finden Sie unter [Authentifizierung und Zugriffskontrolle für AWS Secrets Manager](#) im AWS Secrets Manager -Benutzerhandbuch.

Das Secret in einer Konfigurationsklassifizierung verwenden

Sie können die `EMR.secret@` Anmerkung zu jeder Konfigurationseigenschaft hinzufügen, um anzugeben, dass ihr Schlüssel-Wert-Paar ein ARN für ein in Secrets Manager gespeichertes Geheimnis enthält.

Das folgende Beispiel zeigt, wie ein Geheimnis ARN in einer Konfigurationsklassifizierung angegeben wird:

```
{
```

```

"Classification":"core-site",
"Properties":{
  "presto.s3.access-key":"<sensitive-access-key>",
  "EMR.secret@presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
}
}

```

Wenn Sie Ihren Cluster erstellen und Ihre kommentierte Konfiguration einreichen, EMR validiert Amazon die Konfigurationseigenschaften. Wenn Ihre Konfiguration gültig ist, entfernt Amazon die EMR Anmerkung aus der Konfiguration und ruft das Geheimnis von Secrets Manager ab, um die eigentliche Konfiguration zu erstellen, bevor es auf den Cluster angewendet wird:

```

{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "presto.s3.secret-key":"<my-secret-key-retrieved-from-Secrets-Manager>"
  }
}

```

Wenn Sie eine Aktion aufrufen wie `DescribeCluster`, EMR gibt Amazon die aktuelle Anwendungskonfiguration auf dem Cluster zurück. Wenn eine Eigenschaft der Anwendungskonfiguration so gekennzeichnet ist, dass sie ein Geheimnis enthält ARN, dann enthält die durch den `DescribeCluster` Aufruf zurückgegebene Anwendungskonfiguration den Wert ARN und nicht den geheimen Wert. Dadurch wird sichergestellt, dass der geheime Wert nur auf dem Cluster sichtbar ist:

```

{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}

```

Den Secret-Wert aktualisieren

Amazon EMR ruft den geheimen Wert aus einer mit Anmerkungen versehenen Konfiguration ab, wenn die angehängte Instanzgruppe gestartet, neu konfiguriert oder ihre Größe geändert wird. Sie können Secrets Manager verwenden, um den Wert eines Secrets zu ändern, das in der Konfiguration eines laufenden Clusters verwendet wird. Wenn Sie dies tun, können Sie eine Rekonfigurationsanfrage an jede Instance-Gruppe senden, die den aktualisierten Wert erhalten soll. Weitere Informationen zur Neukonfiguration einer Instance-Gruppe und zu den Dingen, die Sie dabei beachten sollten, finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren

EMRAmazon-Versionen haben unterschiedliche Standardversionen von Java Virtual Machine (JVM). Auf dieser Seite wird die JVM Unterstützung für verschiedene Versionen und Anwendungen erklärt.

Überlegungen

Informationen zu den unterstützten Java-Versionen für Anwendungen finden Sie auf den Anwendungsseiten im [Amazon EMR Release Guide](#).

- Amazon unterstützt EMR nur die Ausführung einer Runtime-Version in einem Cluster und unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Runtime-Versionen auf demselben Cluster.
- Für Amazon EMR 7.x ist Java 17 die standardmäßige Java Virtual Machine (JVM) für Anwendungen, die Java 17 unterstützen, mit Ausnahme von Apache Livy. Weitere Informationen zu den unterstützten JDK Versionen für Anwendungen finden Sie auf der entsprechenden Release-Seite im Amazon EMR Release Guide.
- Ab Amazon EMR 7.1.0 unterstützt Flink Java 17 und ist standardmäßig auf Java 17 eingestellt. Um eine andere Version der Java-Runtime zu verwenden, überschreiben Sie die Einstellungen in `flink-conf`. Weitere Informationen zur Konfiguration von Flink für die Verwendung von Java 8 oder Java 11 finden [Sie unter Konfigurieren von Flink für die Ausführung mit Java 11](#).
- Für Amazon EMR 5.x und 6.x ist Java 8 die standardmäßige Java Virtual Machine (JVM).
 - Für EMR Amazon-Versionen 6.12.0 und höher unterstützen einige Anwendungen auch Java 11 und 17.

- Für EMR Amazon-Versionen 6.9.0 und höher unterstützt Trino standardmäßig Java 17. Weitere Informationen über Java 17 mit Trino finden Sie unter [Trino-Updates für Java 17](#) im Trino-Blog.

Beachten Sie bei der Auswahl Ihrer Laufzeit-Version die folgenden anwendungsspezifischen Überlegungen:

Anwendungsspezifische Java-Konfigurationshinweise

Anwendung	Hinweise zur Java-Konfigurationen
Spark	<p>Um Spark mit einer nicht standardmäßigen Java-Version auszuführen, müssen Sie sowohl Spark als auch Hadoop konfigurieren. Beispiele finden Sie unter Überschreiben Sie die JVM.</p> <ul style="list-style-type: none"> • Konfigurieren Sie <code>JAVA_HOME</code> in <code>spark-env</code>, um die Java-Laufzeit der primären Instanceprozesse zu aktualisieren. Zum Beispiel <code>spark-submit</code>, <code>spark-shell</code> und Spark History Server. • Ändern Sie die Hadoop-Konfiguration, um die Java-Laufzeit der Spark-Executors und der YARN ApplicationMaster
Spark RAPIDS	Sie können RAPIDS mit der konfigurierten Java-Version für Spark arbeiten.
Iceberg	Sie können Iceberg mit der konfigurierten Java-Version der Anwendung ausführen, die es verwendet.
Delta	Sie können Delta mit der konfigurierten Java-Version der Anwendung ausführen, die es verwendet.

Anwendung	Hinweise zur Java-Konfigurationen
Hudi	Sie können Hudi mit der konfigurierten Java-Version der Anwendung ausführen, die es verwendet.
Hadoop	Um das JVM für Hadoop zu aktualisieren, ändern Sie <code>hadoop-env</code> . Beispiele finden Sie unter Überschreiben Sie die JVM .
Hive	Um die Java-Version für Hive auf 11 oder 17 festzulegen, konfigurieren Sie die JVM Hadoop-Einstellung auf die Java-Version, die Sie verwenden möchten.
HBase	Um das JVM für zu aktualisieren HBase, ändern Sie es. <code>hbase-env</code> Standardmäßig EMR legt Amazon das HBase JVM basierend auf der JVM Konfiguration für Hadoop fest, sofern Sie die Einstellungen in <code>hbase-env</code> nicht überschreiben. Beispiele finden Sie unter Überschreiben Sie die JVM .
Flink	Um das JVM für Flink zu aktualisieren, ändern Sie <code>flink-conf</code> Standardmäßig EMR legt Amazon den Flink auf der JVM Grundlage der JVM Konfiguration für Hadoop fest, sofern Sie die Einstellungen in nicht überschreiben. <code>flink-conf</code> Weitere Informationen finden Sie unter Flink für die Ausführung mit Java 11 konfigurieren .

Anwendung	Hinweise zur Java-Konfigurationen
Oozie	Um Oozie für die Ausführung unter Java 11 oder 17 zu konfigurieren, konfigurieren Sie Oozie Server, den Oozie Launcher AM und ändern Sie Ihre clientseitigen ausführbaren Dateien und Auftragskonfigurationen. Sie können <code>EmbeddedOozieServer</code> auch für die Ausführung unter Java 17 konfigurieren. Weitere Informationen finden Sie unter Die Java-Version für Oozie konfigurieren .
Pig	Pig unterstützt nur Java 8. Sie können Java 11 oder 17 nicht mit Hadoop verwenden und Pig auf demselben Cluster ausführen.

Überschreiben Sie die JVM

Um die JVM Einstellung für eine EMR Amazon-Version zu überschreiben — zum Beispiel um Java 17 mit einem Cluster zu verwenden, der Amazon EMR Version 6.12.0 verwendet — geben Sie die `JAVA_HOME` Einstellung an die Umgebungsklassifizierung an, die *application*-env für alle Anwendungen außer Flink gilt. Für Flink lautet die Umgebungsklassifizierung `flink-conf`. Schritte zur Konfiguration der Java-Laufzeit mit Flink finden Sie unter [Flink für die Ausführung mit Java 11 konfigurieren](#).

Themen

- [Überschreiben Sie die JVM Einstellung mit Apache Spark](#)
- [Überschreiben Sie die JVM Einstellung mit Apache HBase](#)
- [Überschreiben Sie die JVM Einstellung mit Apache Hadoop und Hive](#)

Überschreiben Sie die JVM Einstellung mit Apache Spark

Wenn Sie Spark mit EMR Amazon-Versionen 6.12 und höher verwenden und einen Treiber für die Einreichung im Cluster-Modus schreiben, verwendet der Treiber Java 8, aber Sie können die Umgebung so einstellen, dass die Executoren Java 11 oder 17 verwenden. Und wenn Sie Spark mit EMR Amazon-Versionen unter 5.x verwenden und einen Treiber für die Einreichung im Cluster-

Modus schreiben, verwendet der Treiber Java 7. Die Umgebung kann dann aber so eingestellt werden, dass die Executors Java 8 verwenden.

Um den JVM für Spark zu überschreiben, empfehlen wir, dass Sie sowohl die Hadoop- als auch die Spark-Klassifizierung festlegen.

```
{
  "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": {}
    }
  ]
}
```

Überschreiben Sie die JVM Einstellung mit Apache HBase

Um die Verwendung von Java 11 HBase zu konfigurieren, können Sie beim Starten des Clusters die folgende Konfiguration festlegen.

```
[
  {
    "Classification": "hbase-env",
```

```
"Configurations": [  
  {  
    "Classification": "export",  
    "Configurations": [],  
    "Properties": {  
      "JAVA_HOME": "/usr/lib/jvm/jre-11"  
    }  
  }  
],  
"Properties": {}  
}
```

Überschreiben Sie die JVM Einstellung mit Apache Hadoop und Hive

Das folgende Beispiel zeigt, wie Sie die für Hadoop und Hive JVM auf Version 17 setzen.

```
[  
  {  
    "Classification": "hadoop-env",  
    "Configurations": [  
      {  
        "Classification": "export",  
        "Configurations": [],  
        "Properties": {  
          "JAVA_HOME": "/usr/lib/jvm/jre-17"  
        }  
      }  
    ],  
    "Properties": {}  
  }  
]
```

Service-Ports

Im Folgenden sind YARN HDFS Service-Ports aufgeführt. Diese Einstellungen stellen die Hadoop-Standardwerte dar. Andere Anwendungsservices werden bei Standard-Ports gehostet, sofern nicht anders dokumentiert. Weitere Informationen enthält die Projektdokumentation der Anwendung.

Porteinstellungen für YARN und HDFS

Einstellung	Hostname/Port
<code>fs.default.name</code>	Standard (<code>hdfs://<i>emrDeterminedIP</i>:8020</code>)
<code>dfs.datanode.address</code>	Standard (<code>0.0.0.0:50010</code>)
<code>dfs.datanode.http.address</code>	Standard (<code>0.0.0.0:50075</code>)
<code>dfs.datanode.https.address</code>	Standard (<code>0.0.0.0:50475</code>)
<code>dfs.datanode.ipc.address</code>	Standard (<code>0.0.0.0:50020</code>)
<code>dfs.http.address</code>	Standard (<code>0.0.0.0:50070</code>)
<code>dfs.https.address</code>	Standard (<code>0.0.0.0:50470</code>)
<code>dfs.secondary.http.address</code>	Standard (<code>0.0.0.0:50090</code>)
<code>yarn.nodemanager.address</code>	Standard (<code>\${yarn.nodemanager.hostname}:0</code>)
<code>yarn.nodemanager.localizer.address</code>	Standard (<code>\${yarn.nodemanager.hostname}:8040</code>)
<code>yarn.nodemanager.webapp.address</code>	Standard (<code>\${yarn.nodemanager.hostname}:8042</code>)
<code>yarn.resourcemanager.address</code>	Standard (<code>\${yarn.resourcemanager.hostname}:8032</code>)
<code>yarn.resourcemanager.admin.address</code>	Standard (<code>\${yarn.resourcemanager.hostname}:8033</code>)
<code>yarn.resourcemanager.resource-tracker.address</code>	Standard (<code>\${yarn.resourcemanager.hostname}:8031</code>)
<code>yarn.resourcemanager.scheduler.address</code>	Standard (<code>\${yarn.resourcemanager.hostname}:8030</code>)

Einstellung	Hostname/Port
<code>yarn.resourcemanager.webapp.address</code>	Standard (<code>\${yarn.resourcemanager.hostname}:8088</code>)
<code>yarn.web-proxy.address</code>	Standard (no-value)
<code>yarn.resourcemanager.hostname</code>	<i>emrDeterminedIP</i>

Note

Der Begriff *emrDeterminedIP* ist eine IP-Adresse, die von der EMR Amazon-Steuerebene generiert wird. In der neueren Version wurde diese Konvention entfernt, mit Ausnahme der Einstellungen `yarn.resourcemanager.hostname` und `fs.default.name`.

Anwendungsbenutzer

Die Anwendungen führen Prozesse als deren eigene Benutzer aus. Zum Beispiel Hive „Als Benutzer JVMs ausführen“ `hivemapred`, „MapReduce JVMs ausführen als“ usw. Dies wird im folgenden Prozessstatusbeispiel gezeigt.

```

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/l
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
```

Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys

Sie können das Amazon-EMR-Artefakt-Repository verwenden, um Apache-Hive- und Apache-Hadoop-Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die mit bestimmten Amazon-EMR-Release-Versionen verfügbar sind, beginnend mit Amazon-EMR-Version 5.18.0. Die Erstellung auf der Grundlage von Amazon-EMR-Artefakten im Repository hilft dabei, Laufzeit-Klassen-Probleme zu vermeiden, da sichergestellt wird, dass die Versionen der Bibliotheken, auf deren Grundlage der Job erstellt wird, exakt den Versionen entsprechen, die zur Laufzeit auf dem Cluster bereitgestellt werden. Derzeit sind Amazon-EMR-Artefakte nur für Maven-Builds verfügbar.

Fügen Sie zum Zugriff auf das Artefakt-Repository Ihrer Maven-Einstellungsdatei oder der `pom.xml`-Konfigurationsdatei eines bestimmten Projekts die Repository-URL hinzu. Anschließend können Sie die Abhängigkeiten in Ihrer Projektkonfiguration angeben. Verwenden Sie für Abhängigkeitsversionen die unter Komponentenversionen für die gewünschte Version auf [Amazon EMR 5.x-Release-Versionen](#) aufgeführte Version. Beispielsweise sind Komponentenversionen für die aktuelle Amazon-EMR-Version unter [the section called "5.36.2 Komponentenversionen"](#) verfügbar. Wenn ein Artefakt für Ihr Projekt nicht unter Komponentenversionen aufgeführt ist, geben Sie die Version an, die in diesem Release für Hive und Hadoop aufgeführt ist. Beispielsweise ist für Hadoop-Komponenten in Amazon-EMR-Version 5.18.0 die Version `2.8.4-amzn-1`.

Die Artefakt-Repository-URL hat die folgende Syntax:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- *s3-endpoint* ist der Amazon Simple Storage Service (Amazon S3)-Endpunkt der Region für das Repository, und *region-ID* ist die entsprechende Region. Beispiel: `s3.us-west-1.amazonaws.com` und `us-west-1`. Weitere Informationen finden Sie unter Endpunkte für Amazon S3 in Allgemeine Amazon Web Services-Referenz. Es gibt keinen Unterschied hinsichtlich der Artefakte zwischen Regionen. So können Sie die am besten geeignete Region für Ihre Entwicklungsumgebung angeben.
- *emr-release-label* ist das Release-Label für den Amazon EMR-Cluster, auf dem Ihr Code ausgeführt wird. Versionsbezeichnungen haben das Format `emr-x.x.x`, z. B. `emr-5.36.2`. Eine EMR-Veröffentlichungsserie kann mehrere Veröffentlichungen enthalten. Wenn Sie beispielsweise

EMR-Version 5.24.1 verwenden, verwenden Sie die erste EMR-Versionsbezeichnung innerhalb der 5.24-Serie, `emr-5.24.0`, in der Artefakt-Repository-URL:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

Example Konfiguration für Maven pom.xml

Das pom.xml-Beispiel unten konfiguriert ein Maven-Projekt zum Erstellen auf der Grundlage der `emr-5.18.0` Apache Hadoop- und Apache Hive-Artefakte unter Verwendung des Artefakt-Repositorys in `us-west-1`. Snapshot-Versionen sind im Artefakt-Repository nicht verfügbar, so dass Snapshots im pom.xml deaktiviert sind. Auslassungspunkte (`...`) im folgenden Beispiel zeigen die Auslassung anderer Konfigurationsparameter an. Kopieren Sie diese nicht in Ihr Maven-Projekt.

```
<project>
  ...
  <repositories>
    ...
    <repository>
      <id>emr-5.18.0-artifacts</id>
      <name>EMR 5.18.0 Releases Repository</name>
      <releases>
        <enabled>true</enabled>
      </releases>
      <snapshots>
        <enabled>false</enabled>
      </snapshots>
      <url>https://s3.us-west-1.amazonaws.com/us-west-1-emr-artifacts/emr-5.18.0/repos/
maven/</url>
    </repository>
    ...
  </repositories>
  ...
  <dependencies>
    ...
    <dependency>
      <groupId>org.apache.hive</groupId>
      <artifactId>hive-exec</artifactId>
      <version>2.3.3-amzn-2</version>
    </dependency>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
```



```
<artifactId>hadoop-common</artifactId>  
<version>2.8.4-amzn-1</version>  
</dependency>  
...  
</dependencies>  
  
</project>
```

EMR Dateisystem (EMRFS)

Das EMR Dateisystem (EMRFS) ist eine Implementierung von HDFS, die alle EMR Amazon-Cluster verwenden, um reguläre Dateien von Amazon EMR direkt in Amazon S3 zu lesen und zu schreiben. EMRFS bietet den Komfort, persistente Daten in Amazon S3 zur Verwendung mit Hadoop zu speichern und bietet gleichzeitig Funktionen wie Datenverschlüsselung.

Datenverschlüsselung ermöglicht es Ihnen, Objekte zu verschlüsseln in EMRFS, die in Amazon S3 schreiben, und ermöglicht EMRFS die Arbeit mit verschlüsselten Objekten in Amazon S3. Wenn Sie die EMR Amazon-Release-Version 4.8.0 oder höher verwenden, können Sie Sicherheitskonfigurationen verwenden, um die Verschlüsselung für EMRFS Objekte in Amazon S3 zusammen mit anderen Verschlüsselungseinstellungen einzurichten. Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#). Wenn Sie eine frühere Version von Amazon verwenden EMR, können Sie die Verschlüsselungseinstellungen manuell konfigurieren. Weitere Informationen finden Sie unter [Amazon S3 S3-Verschlüsselung mithilfe von EMRFS Eigenschaften angeben](#).

Amazon S3 bietet eine hohe Konsistenz beim Lesen nach dem Schreiben für alle GET/PUT, und alle LIST Operationen. AWS-Regionen Das bedeutet, dass das, was Sie verwenden, das EMRFS ist, was Sie aus Amazon S3 lesen, ohne dass die Leistung beeinträchtigt wird. Weitere Informationen finden Sie unter [Amazon-S3-Datenkonsistenzmodell](#).

Wenn Sie die EMR Amazon-Release-Version 5.10.0 oder höher verwenden, können Sie je nach Cluster-Benutzern, Gruppen oder dem Speicherort der EMRFS Daten in Amazon S3 unterschiedliche IAM Rollen für EMRFS Anfragen an Amazon S3 verwenden. Weitere Informationen finden [Sie unter IAM Rollen für EMRFS Anfragen an Amazon S3 konfigurieren](#).

Warning

Bevor Sie die spekulative Ausführung für EMR Amazon-Cluster aktivieren, auf denen Apache Spark-Jobs ausgeführt werden, lesen Sie bitte die folgenden Informationen.

EMRFS beinhaltet den EMRFS S3-optimierten Committer, eine OutputCommitter Implementierung, die für das Schreiben von Dateien in Amazon S3 bei Verwendung optimiert ist. EMRFS [Wenn Sie die spekulative Ausführungsfunktion von Apache Spark mit Anwendungen aktivieren, die Daten in Amazon S3 schreiben und den EMRFS S3-optimierten Committer nicht verwenden, können Probleme mit der Datenkorrektheit auftreten, die in -10063 beschrieben werden. SPARK](#) Dies kann auftreten, wenn Sie EMR Amazon-Versionen vor EMR Amazon-Version 5.19 verwenden oder wenn Sie Dateien mit Formaten wie ORC und CSV auf Amazon S3 schreiben. Diese Formate werden vom EMRFS S3-

optimierten Committer nicht unterstützt. [Eine vollständige Liste der Anforderungen für die Verwendung des S3-optimierten Committers finden Sie unter Anforderungen für den EMRFS S3-optimierten Committer. EMRFS](#)

EMRFS Direct Write wird normalerweise verwendet, wenn der EMRFS S3-optimierte Committer nicht unterstützt wird, z. B. wenn Folgendes geschrieben wird:

- Ein anderes Ausgabeformat als Parquet, z. B. ORC oder Text.
- Hadoop-Dateien, die den Spark RDD API verwenden.
- Parquet mit Hive SerDe. Sehen Sie unter [Konvertierung der Parquet-Tabelle in Hive Metastore](#).

EMRFS Direct Write wird in den folgenden Szenarien nicht verwendet:

- Wenn der EMRFS S3-optimierte Committer aktiviert ist. Weitere Informationen finden Sie unter [Anforderungen für den S3-optimierten Committer EMRFS](#).
- Beim Schreiben dynamischer Partitionen mit der `partitionOverwriteMode` Einstellung „dynamisch“.
- Beim Schreiben in benutzerdefinierte Partitionsspeicherorte, z. B. an Speicherorte, die nicht der Hive-Standardkonvention für Partitionen entsprechen.
- Wenn Sie andere Dateisysteme als das S3A-Dateisystem verwenden EMRFS, z. B. wenn Sie in das S3A-Dateisystem schreiben HDFS oder es verwenden.

Um festzustellen, ob Ihre Anwendung Direct Write in Amazon EMR 5.14.0 oder höher verwendet, aktivieren Sie die INFO Spark-Protokollierung. Wenn eine Protokollzeile mit dem Text „Direct Write:ENABLED“ entweder in den Spark-Treiberprotokollen oder den Spark-Executor-Container-Protokollen vorhanden ist, dann hat Ihre Spark-Anwendung mit Direct Write geschrieben.

Standardmäßig ist die spekulative Ausführung bei Amazon OFF EMRclusters aktiviert. Wir empfehlen dringend, die spekulative Ausführung nicht zu aktivieren, wenn diese beiden Bedingungen zutreffen:

- Sie schreiben Daten in Amazon S3.
- Daten werden in einem anderen Format als Apache Parquet oder im Apache Parquet-Format ohne Verwendung des EMRFS S3-optimierten Committers geschrieben.

Wenn Sie die spekulative Ausführung von Spark aktivieren und Daten mit EMRFS Direct Write auf Amazon S3 schreiben, kann es zu zeitweiligem Datenverlust kommen. Wenn Sie mit dem EMRFS S3-optimierten Committer Daten in Parquet schreiben oder Daten in Parquet schreiben, verwendet Amazon EMR kein direktes Schreiben und dieses Problem tritt nicht auf. HDFS

Wenn Sie Daten in Formaten schreiben müssen, die EMRFS direktes Schreiben von Spark nach Amazon S3 verwenden und spekulative Ausführung verwenden, empfehlen wir, mit S3DistCP in Amazon S3 zu schreiben HDFS und dann Ausgabedateien an Amazon S3 zu übertragen.

Themen

- [Consistent view](#)
- [Autorisieren des Zugriffs auf EMRFS Daten in Amazon S3](#)
- [Den AWS Security Token Service Standardendpunkt verwalten](#)
- [Amazon S3 S3-Verschlüsselung mithilfe von EMRFS Eigenschaften angeben](#)

Consistent view

Warning

Am 1. Juni 2023 wird EMRFS Consistent View das Ende der Standardunterstützung für future EMR Amazon-Versionen erreichen. EMRFS Die konsistente Ansicht wird auch weiterhin für bestehende Versionen funktionieren.

Mit der Veröffentlichung von Amazon S3 Strong read-after-write Consistency am 1. Dezember 2020 müssen Sie EMRFS Consistent View (EMRFSCV) nicht mehr für Ihre EMR Amazon-Cluster verwenden. EMRFSCV ist eine optionale Funktion, mit der EMR Amazon-Cluster Amazon S3-Objekte auf Liste und read-after-write Konsistenz überprüfen können. Wenn Sie einen Cluster erstellen und EMRFS CV aktiviert ist, EMR erstellt Amazon eine Amazon DynamoDB DynamoDB-Datenbank zum Speichern von Objektmetadaten, die zur Nachverfolgung und read-after-write Konsistenz von S3-Objekten verwendet werden. Sie können EMRFS CV jetzt deaktivieren und die DynamoDB DynamoDB-Datenbank löschen, sodass Ihnen keine zusätzlichen Kosten entstehen.

In den folgenden Verfahren wird erklärt, wie Sie das CV-Feature prüfen, ausschalten und die DynamoDB-Datenbank löschen, die das Feature verwendet.

Um zu überprüfen, ob Sie die CV-Funktion verwenden EMRFS

1. Navigieren Sie zur Registerkarte Konfiguration. Wenn Ihr Cluster die folgende Konfiguration hat, verwendet er EMRFS CV.

```
Classification=emrfs-site,Property=fs.s3.consistent,Value=true
```

2. Verwenden Sie alternativ den, AWS CLI um Ihren Cluster mit dem zu beschreiben [describe-clusterAPI](#). Wenn die Ausgabe Folgendes enthält `fs.s3.consistent: true`, verwendet Ihr Cluster EMRFS CV.

So schalten Sie EMRFS CV auf Ihren EMR Amazon-Clustern aus

Verwenden Sie eine der folgenden drei Optionen, um die EMRFS CV-Funktion zu deaktivieren. Sie sollten diese Optionen in Ihrer Testumgebung testen, bevor Sie sie auf Ihre Produktionsumgebungen anwenden.

1. Um Ihren vorhandenen Cluster zu beenden und einen neuen Cluster ohne EMRFS CV-Optionen zu starten.
 - a. Bevor Sie Ihren Cluster beenden, stellen Sie sicher, dass Sie Ihre Daten sichern und Ihre Benutzer benachrichtigen.
 - b. Befolgen Sie die Anweisungen unter [Einen Cluster beenden](#), um den Cluster zu beenden.
 - c. Wenn Sie die EMR Amazon-Konsole verwenden, um einen neuen Cluster zu erstellen, navigieren Sie zu Erweiterte Optionen. Deaktivieren Sie im Abschnitt Softwareeinstellungen bearbeiten die Option zum Aktivieren von EMRFS CV. Wenn das Kontrollkästchen für eine EMRFS einheitliche Ansicht verfügbar ist, lassen Sie es deaktiviert.
 - d. Wenn Sie früher AWS CLI einen neuen Cluster mit dem erstellen [create-clusterAPI](#), verwenden Sie nicht die `--emrfs` Option, die EMRFS CV aktiviert.
 - e. Wenn Sie ein SDK oder verwenden, AWS CloudFormation um einen neuen Cluster zu erstellen, verwenden Sie keine der Konfigurationen, die [unter Konsistente Ansicht konfigurieren](#) aufgeführt sind.
2. Um einen Cluster zu klonen und EMRFS CV zu entfernen
 - a. Wählen Sie in der EMR Amazon-Konsole den Cluster aus, der EMRFS CV verwendet.

- b. Wählen Sie oben auf der Seite Cluster Details (Clusterdetails) auf Clone (Klonen).
- c. Wählen Sie Zurück und navigieren Sie zu Schritt 1: Software und Schritte.
- d. Entfernen Sie unter Softwareeinstellungen bearbeiten die Option EMRFS CV. Löschen Sie unter Konfiguration bearbeiten die folgenden Konfigurationen in der `emrfs-site`-Klassifizierung. Wenn Sie JSON aus einem S3-Bucket laden, müssen Sie Ihr S3-Objekt ändern.

```
[
  {"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. Um EMRFS CV aus einem Cluster zu entfernen, der Instanzgruppen verwendet
 - a. Verwenden Sie den folgenden Befehl, um zu überprüfen, ob ein einzelner EMR Cluster die DynamoDB-Tabelle verwendet, die mit EMRFS CV verknüpft ist, oder ob mehrere Cluster die Tabelle gemeinsam nutzen. Der Tabellename ist in `fs.s3.consistent.metadata.tableName` angegeben, wie unter [Konsistente Ansicht konfigurieren beschrieben](#). Der von EMRFS CV verwendete Standardtabellename ist `EmrFSMetadata`

```
aws emr describe-cluster --cluster-id j-XXXXX | grep
fs.s3.consistent.metadata.tableName
```

- b. Wenn Ihr Cluster Ihre DynamoDB-Datenbank nicht mit einem anderen Cluster teilt, verwenden Sie den folgenden Befehl, um den Cluster neu zu konfigurieren und CV zu deaktivieren. EMRFS Weitere Informationen finden Sie unter [Neu-Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

```
aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json
```

Mit diesem Befehl wird die Datei geöffnet, die Sie ändern möchten. Ändern Sie die Datei mit den folgenden Konfigurationen.

```
{
  "ClusterId": "j-xxxx",
  "InstanceGroups": [
    {
      "InstanceGroupId": "ig-xxxx",
      "Configurations": [
        {
          "Classification": "emrfs-site",
          "Properties": {
            "fs.s3.consistent": "false"
          },
          "Configurations": []
        }
      ]
    }
  ]
}
```

- c. Wenn Ihr Cluster die DynamoDB-Tabelle mit einem anderen Cluster teilt, deaktivieren Sie EMRFS CV auf allen Clustern zu einem Zeitpunkt, zu dem keine Cluster Objekte am gemeinsam genutzten S3-Standort ändern.

So löschen Sie Amazon DynamoDB DynamoDB-Ressourcen, die mit CV verknüpft sind EMRFS

Nachdem Sie EMRFS CV aus Ihren EMR Amazon-Clustern entfernt haben, löschen Sie die mit EMRFS CV verknüpften DynamoDB-Ressourcen. Bis Sie dies tun, fallen für Sie weiterhin DynamoDB-Gebühren im Zusammenhang mit CV an. EMRFS

1. Überprüfen Sie die CloudWatch Metriken für Ihre DynamoDB-Tabelle und stellen Sie sicher, dass die Tabelle von keinem Cluster verwendet wird.
2. Löschen Sie die DynamoDB-Tabelle.

```
aws dynamodb delete-table --table-name <your-table-name>
```

So löschen Sie SQS Amazon-Ressourcen, die mit dem EMRFS Lebenslauf verknüpft sind

1. Wenn Sie Ihren Cluster so konfiguriert haben, dass Inkonsistenzbenachrichtigungen an Amazon gesendet werden, können Sie alle SQS Warteschlangen löschen.
2. Suchen Sie den in angegebenen SQS Amazon-Warteschlangennamen `s3.consistent.notification.SQS.queueName`, wie unter [Konsistente Ansicht konfigurieren](#) beschrieben. Das Standardformat für Warteschlangennamen ist `EMRFS-Inconsistency-<j-cluster ID>`.

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'  
aws sqs delete-queue --queue-url <your-queue-url>
```

Um die Verwendung des zu beenden EMRFS CLI

- Der [EMRFSCLI](#) verwaltet die Metadaten, die EMRFS CV generiert. Da der Standardsupport für EMRFS CV in future Versionen von Amazon EMR ausläuft, wird auch der Support für den eingestellt.

Themen

- [Die konsistente Ansicht aktivieren](#)
- [Verstehen, wie EMRFS Consistent View Objekte in Amazon S3 verfolgt](#)
- [Logik für Wiederholversuche](#)
- [EMRFS konsistente Metadaten anzeigen](#)
- [Konsistenzbenachrichtigungen für CloudWatch und Amazon konfigurieren SQS](#)
- [Konfigurieren der konsistenten Ansicht](#)
- [EMRFSCLIBefehlsreferenz](#)

Die konsistente Ansicht aktivieren

Sie können die serverseitige Amazon S3 S3-Verschlüsselung oder die konsistente Ansicht für die EMRFS Verwendung der AWS Management Console AWS CLI, oder der `emrfs-site` Konfigurationsklassifizierung aktivieren.

So konfigurieren Sie die konsistente Ansicht mit der Konsole

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie in Step 1: Software and Steps (Schritt 1: Software und Schritte) und Step 2: Hardware (Schritt 2: Hardware) Einstellungen aus.
4. Wählen Sie für Schritt 3: Allgemeine Cluster-Einstellungen unter Zusätzliche Optionen die Option EMRFS Konsistente Ansicht aus.
5. Geben Sie EMRFS unter Metadatenpeicher den Namen Ihres MetadatenSpeichers ein. Der Standardwert ist **EmrFSMetadata**. Wenn die mrFSMetadata E-Tabelle nicht existiert, wird sie in DynamoDB für Sie erstellt.

Note

Amazon entfernt die EMRFS Metadaten EMR nicht automatisch aus DynamoDB, wenn der Cluster beendet wird.

6. Geben Sie in Number of retries (Anzahl der Wiederholungen) einen Ganzzahlwert ein. Wenn eine Inkonsistenz festgestellt wird, EMRFS versucht Amazon S3 so oft aufzurufen. Der Standardwert ist **5**.
7. Geben Sie in Retry period (in seconds) (Wiederholungszeitraum (in Sekunden)) einen Ganzzahlwert ein. Dies ist die Zeitspanne, die zwischen EMRFS Wiederholungsversuchen vergeht. Der Standardwert ist **10**.

Note

Bei nachfolgenden Wiederholversuchen wird ein exponentielles Backoff verwendet.

Um einen Cluster mit aktivierter konsistenter Ansicht zu starten, verwenden Sie AWS CLI

Wir empfehlen Ihnen, die aktuelle Version von zu installieren AWS CLI. Informationen zum Herunterladen der neuesten Version finden Sie unter <https://aws.amazon.com/cli/>.

• **Note**

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs
  Consistent=true \
--release-label emr-7.2.0 --ec2-attributes KeyName=myKey
```

Um zu überprüfen, ob die konsistente Ansicht aktiviert ist, verwenden Sie AWS Management Console

- Um in der Konsole zu überprüfen, ob die konsistente Ansicht aktiviert ist, navigieren Sie zu Cluster List (Cluster-Liste) und wählen den Namen Ihres Clusters aus, um die Cluster Details (Cluster-Details) anzuzeigen. Das Feld „EMRFSKonsistente Ansicht“ hat den Wert `Enabled` oder `Disabled`.

So prüfen Sie durch Untersuchen der Datei `emrfs-site.xml`, ob die konsistente Ansicht aktiviert ist

- Sie können überprüfen, ob die Konsistenz aktiviert ist, indem Sie die Konfigurationsdatei `emrfs-site.xml` auf dem Master-Knoten des Clusters prüfen. Wenn der boolesche Wert für `fs.s3.consistent` auf `true` festgelegt ist, ist die konsistente Ansicht für Dateisystem-Operationen, an denen Amazon S3 beteiligt ist, aktiviert.

Verstehen, wie EMRFS Consistent View Objekte in Amazon S3 verfolgt

EMRFS erstellt eine konsistente Ansicht von Objekten in Amazon S3, indem Informationen über diese Objekte zu den EMRFS Metadaten hinzugefügt werden. EMRFS fügt diese Auflistungen zu seinen Metadaten hinzu, wenn:

- Ein Objekt, das von EMRFS im Laufe eines EMR Amazon-Jobs geschrieben wurde.
- Ein Objekt wird mithilfe von mit Metadaten synchronisiert oder in EMRFS Metadaten importiert.

EMRFS CLI

Objekte, von EMRFS denen gelesen wird, werden den Metadaten nicht automatisch hinzugefügt. Wenn ein Objekt EMRFS gelöscht wird, verbleibt eine Liste in den Metadaten mit dem Status Gelöscht, bis diese Liste mithilfe von gelöscht wird. EMRFS CLI Weitere Informationen über das finden Sie unter CLI. [EMRFSCLIBefehlsreferenz](#) Weitere Informationen zum Löschen von Auflistungen in den EMRFS Metadaten finden Sie unter [EMRFSkonsistente Metadaten anzeigen](#).

EMRFS überprüft bei jedem Amazon S3 S3-Vorgang die Metadaten auf Informationen über die Gruppe von Objekten in einer konsistenten Ansicht. Wenn EMRFS festgestellt wird, dass Amazon S3 während eines dieser Vorgänge inkonsistent ist, wiederholt es den Vorgang gemäß den in den `emrfs-site` Konfigurationseigenschaften definierten Parametern. Wenn die Wiederholungsversuche EMRFS erschöpft sind, wird entweder eine Ausnahme ausgelöst `ConsistencyException` oder die Ausnahme protokolliert und der Workflow wird fortgesetzt. Weitere Informationen zur Wiederholungslogik finden Sie unter [Logik für Wiederholversuche](#). Sie finden die `ConsistencyExceptions` in Ihren Protokollen, zum Beispiel:

- `listStatus`: Kein Amazon S3 S3-Objekt für Metadatenelement `/S3_bucket/dir/object`
- `getFileStatus`: Der Schlüssel `dir/file` ist in den Metadaten vorhanden, aber nicht in Amazon S3

Wenn Sie ein Objekt direkt aus Amazon S3 löschen, das EMRFS Consistent View verfolgt, EMRFS behandelt dieses Objekt als inkonsistent, da es in den Metadaten immer noch als in Amazon S3 vorhanden aufgeführt ist. Wenn Ihre Metadaten nicht mehr mit den EMRFS Objektspuren in Amazon S3 synchron sind, können Sie den `sync` Unterbefehl von verwenden, EMRFS CLI um Metadaten so zurückzusetzen, dass sie Amazon S3 entsprechen. Um Abweichungen zwischen Metadaten und Amazon S3 zu entdecken, verwenden Sie den Befehl `diff`. Schließlich bietet es EMRFS nur eine konsistente Ansicht der Objekte, auf die in den Metadaten verwiesen wird. Es kann andere Objekte im selben Amazon S3-Pfad geben, die nicht verfolgt werden. Wenn die Objekte in einem Amazon S3 S3-Pfad EMRFS aufgelistet werden, wird die Obermenge der Objekte zurückgegeben, die in den Metadaten nachverfolgt werden, und der Objekte in diesem Amazon S3 S3-Pfad.

Logik für Wiederholversuche

EMRFS versucht, die Listenkonsistenz für Objekte, die in den Metadaten erfasst werden, für eine bestimmte Anzahl von Wiederholungen zu überprüfen. Der Standardwert ist 5. Wenn die Anzahl der Wiederholversuche überschritten wird, gibt der ursprüngliche Auftrag einen Fehler zurück, es sei denn, `fs.s3.consistent.throwExceptionOnInconsistency` ist auf `false` festgelegt, sodass die Objekte lediglich protokolliert werden, die als inkonsistent verfolgt wurden. EMRFS verwendet standardmäßig eine exponentielle Backoff-Wiederholungsrichtlinie,

aber Sie können sie auch auf eine feste Richtlinie festlegen. Möglicherweise möchten die Benutzer ihre Versuche für einen bestimmten Zeitraum wiederholen, bevor sie den restlichen Auftrag ausführen, ohne dass eine Ausnahme ausgelöst wird. Sie können dies erreichen, indem sie `fs.s3.consistent.throwExceptionOnInconsistency` auf `false`, `fs.s3.consistent.retryPolicyType` auf `fixed` und `fs.s3.consistent.retryPeriodSeconds` auf den gewünschten Wert festlegen. Im folgenden Beispiel wird ein Cluster mit aktivierter Konsistenz erstellt. Dabei werden Inkonsistenzen protokolliert und es wird ein festes Wiederholungsintervall von 10 Sekunden festgelegt:

Example Festlegen des Wiederholungszeitraums auf einen festen Wert

```
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

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

Weitere Informationen finden Sie unter [Consistent view](#).

EMRFS Konfigurationen für Anrufe nach Regionen abrufen IMDS

EMRFSverlässt sich auf den IMDS (Instance-Metadaten-Service), um die Instance-Region und Amazon S3, DynamoDB oder AWS KMS Endpunkte abzurufen. Es gibt jedoch ein Limit für die Anzahl der Anfragen, die es verarbeiten kann, und Anfragen, die dieses Limit überschreiten, schlagen fehl. Dieses IMDS Limit kann dazu führen, dass die Initialisierung EMRFS fehlschlägt und die Abfrage oder der Befehl fehlschlagen. Sie können den folgenden Mechanismus für zufällige exponentielle Backoff-Wiederholungen und die Konfigurationseigenschaften einer Fallback-Region in `emrfs-site.xml` verwenden, um das Szenario zu beheben, in dem alle Wiederholungsversuche fehlschlagen.

```
<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>
```

EMRFSkonsistente Metadaten anzeigen

EMRFS Die konsistente Ansicht verfolgt die Konsistenz mithilfe einer DynamoDB-Tabelle, um Objekte in Amazon S3 zu verfolgen, die mit Amazon S3 synchronisiert oder von erstellt wurden. EMRFS Mithilfe dieser Metadaten werden alle Operationen (Lesen, Schreiben, Aktualisieren und Kopieren) verfolgt. Sie enthalten keine tatsächlichen Inhalte. Anhand der Metadaten wird überprüft, ob die Objekte oder Metadaten, die von Amazon S3 eingehen, den erwarteten Werten entsprechen. Diese Bestätigung bietet EMRFS die Möglichkeit, die Konsistenz und read-after-write Konsistenz der Listen für neue Objekte, die in Amazon S3 EMRFS geschrieben wurden, oder für Objekte, mit EMRFS denen synchronisiert wurde, zu überprüfen. Mehrere Cluster können dieselben Metadaten verwenden.

So fügen Sie Einträge zu den Metadaten hinzu

Sie können die Unterbefehle `import` oder `sync` verwenden, um Metadaten Einträge hinzuzufügen. `sync` gibt den Status der Amazon-S3-Objekte in einem Pfad an, während `import` ausschließlich zum Hinzufügen neuer Einträge zu den Metadaten verwendet wird. Weitere Informationen finden Sie unter [EMRFSCLIBefehlsreferenz](#).

So überprüfen Sie Unterschiede zwischen Metadaten und Objekten in Amazon S3

Um nach Unterschieden zwischen den Metadaten und Amazon S3 zu suchen, verwenden Sie den `diff` Unterbefehl von EMRFS CLI. Weitere Informationen finden Sie unter [EMRFSCLIBefehlsreferenz](#).

So erkennen Sie, ob Metadatenoperationen gedrosselt werden

EMRFS legt die standardmäßigen Durchsatzkapazitätsgrenzen für die Metadaten für ihre Lese- und Schreibvorgänge auf 500 bzw. 100 Einheiten fest. Eine große Anzahl von Objekten oder Buckets kann dazu führen, dass Vorgänge dieses Limit überschreiten. Sie werden dann von DynamoDB gedrosselt. Beispielsweise kann eine Anwendung EMRFS dazu führen, dass ein ausgelöst wird, `ProvisionedThroughputExceededException` wenn Sie einen Vorgang ausführen, der diese Kapazitätsgrenzen überschreitet. Bei der Drosselung versucht das EMRFS CLI Tool, mithilfe des [exponentiellen Backoffs](#) erneut zu versuchen, in die DynamoDB-Tabelle zu schreiben, bis der Vorgang abgeschlossen ist oder wenn der maximale Wiederholungswert für das Schreiben von Objekten von Amazon nach Amazon S3 erreicht ist. EMR

Sie können Ihre eigene Durchsatzkapazitätsgrenzen konfigurieren. DynamoDB hat jedoch strenge Partitions Grenzen von 3000 Lesekapazitätseinheiten (RCUs) und 1000 Schreibkapazitätseinheiten (WCUs) pro Sekunde für Lese- und Schreiboperationen. Um durch Drosselung verursachte sync Ausfälle zu vermeiden, empfehlen wir, den Durchsatz für Lesevorgänge auf weniger als 3000 RCUs und für Schreibvorgänge auf weniger als 1000 zu begrenzen. WCUs Anweisungen zur Festlegung benutzerdefinierter Durchsatzkapazitätsgrenzen finden Sie unter [Konfigurieren der konsistenten Ansicht](#).

Sie können auch CloudWatch Amazon-Metriken für Ihre EMRFS Metadaten in der DynamoDB-Konsole anzeigen, wo Sie die Anzahl der gedrosselten Lese- und Schreiboperationen sehen können. Wenn für gedrosselte Anforderungen ein Wert ungleich null vorliegt, profitiert Ihre Anwendung möglicherweise von einer Erhöhung der zugewiesenen Durchsatzkapazität für Lese- oder Schreiboperationen. Sie realisieren ggf. auch dann eine höhere Performance, wenn Sie sehen, dass sich Ihre Operationen dem zugewiesenen Höchstwert für die Durchsatzkapazität für Lese- und Schreibvorgänge für einen längeren Zeitraum nähern.

Durchsatzeigenschaften für wichtige Operationen EMRFS

Die Standardeinstellung für Lese- und Schreibvorgänge beträgt 400 bzw. 100 Durchsatzkapazitätseinheiten. Die folgenden Leistungsmerkmale geben Ihnen eine Vorstellung davon, welcher Durchsatz für bestimmte Operationen erforderlich ist. Diese Tests wurden unter Verwendung eines `m3.1large`-Clusters mit einem einzigen Knoten durchgeführt. Bei allen Operationen handelte es sich um Single-Thread-Vorgänge. Die Performance kann je nach

Charakteristika der jeweiligen Anwendung sehr unterschiedlich ausfallen und etwas Experimentieren erfordern, um die Dateisystemoperationen zu optimieren.

Operation	Durchschnittlich read-per-second	Durchschnittlich write-per-second
create (Objekt)	26.79	6.70
delete (Objekt)	10.79	10.79
delete (Verzeichnis mit 1000 Objekten)	21.79	338.40
getFileStatus(Objekt)	34.70	0
getFileStatus(Verzeichnis)	19.96	0
listStatus(Verzeichnis mit 1 Objekt)	43.31	0
listStatus(Verzeichnis mit 10 Objekten)	44.34	0
listStatus(Verzeichnis mit 100 Objekten)	84.44	0
listStatus(Verzeichnis mit 1.000 Objekten)	308.81	0
listStatus(Verzeichnis mit 10.000 Objekten)	416.05	0
listStatus(Verzeichnis mit 100.000 Objekten)	823.56	0
listStatus(Verzeichnis mit 1 Million Objekten)	882.36	0
mkdir (durchgehend für 120 Sekunden)	24.18	4.03

Operation	Durchschnittlich read-per-second	Durchschnittlich write-per-second
mkdir	12.59	0
rename (Objekt)	19.53	4.88
rename (Verzeichnis mit 1000 Objekten)	23.22	339.34

So übermitteln Sie einen Schritt, der alle alten Daten aus Ihren Metadaten entfernt

Unter Umständen möchten Benutzer bestimmte Einträge in den DynamoDB-basierten Metadaten entfernen. So können Sie die Speicherkosten im Zusammenhang mit der Tabelle reduzieren. Benutzer haben die Möglichkeit, bestimmte Einträge mithilfe des Unterbefehls manuell oder programmgesteuert zu löschen. EMRFS CLI `delete` Wenn Sie jedoch Einträge aus den Metadaten löschen, werden EMRFS keine Konsistenzprüfungen mehr durchgeführt.

Das programmgesteuerte Löschen nach Abschluss eines Jobs kann durchgeführt werden, indem Sie einen letzten Schritt an Ihren Cluster senden, der einen Befehl auf dem ausführt. EMRFS CLI Geben Sie z. B. den folgenden Befehl ein, um einen Schritt zum Löschen aller Einträge an Ihren Cluster zu übermitteln, die älter als zwei Tage sind.

```
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"
  ]
}
```

Verwenden Sie den zurückgegebenen `StepId` Wert, um die Protokolle auf das Ergebnis des Vorgangs zu überprüfen.

Konsistenzbenachrichtigungen für CloudWatch und Amazon konfigurieren SQS

Sie können CloudWatch Metriken und SQS Amazon-Nachrichten EMRFS für eventuelle Konsistenzprobleme in Amazon S3 aktivieren.

CloudWatch

Wenn CloudWatch Metriken aktiviert sind, wird jedes Mal, wenn ein `FileSystem` API Anruf aufgrund der eventuellen Konsistenz von Amazon S3 fehlschlägt, eine Metrik namens `Inconsistency` übertragen.

So zeigen Sie CloudWatch Metriken für eventuelle Konsistenzprobleme in Amazon S3 an

Um die Inkonsistenzmetrik in der CloudWatch Konsole anzuzeigen, wählen Sie die EMRFS Metriken aus und wählen Sie dann ein Paar `JobFlowId/Metric Name` aus. Beispiel: `j-162XXXXXXM2CU ListStatus`, `j-162XXXXXXM2CU GetFileStatus` usw.

1. Öffnen Sie die CloudWatch Konsole unter. <https://console.aws.amazon.com/cloudwatch/>
2. Wählen Sie im Dashboard im Bereich Metriken die Option EMRFS.
3. Wählen Sie im Bereich Job Flow Metrics ein oder mehrere `JobFlowId/Metric Name`-Paare aus. Im Fenster darunter wird eine grafische Darstellung der Metriken angezeigt.

Amazon SQS

Wenn SQS Amazon-Benachrichtigungen aktiviert sind, EMRFS-`Inconsistency-<jobFlowId>` wird bei EMRFS der Initialisierung eine SQS Amazon-Warteschlange mit dem Namen erstellt. SQSAmazon-Nachrichten werden in die Warteschlange verschoben, wenn ein `FileSystem` API Anruf aufgrund der eventuellen Konsistenz von Amazon S3 fehlschlägt. Die Nachricht enthält Informationen wie `JobFlowId` eine Liste inkonsistenter Pfade, einen Stack-Trace usw. API Nachrichten können mit der SQS Amazon-Konsole oder mit dem EMRFS `read-sqs` Befehl gelesen werden.

So verwalten Sie SQS Amazon-Nachrichten für Amazon S3, falls Konsistenzprobleme auftreten

Eventuelle Konsistenzprobleme bei SQS Amazon-Nachrichten für Amazon S3 können mit dem EMRFS CLI gelesen werden. Um Nachrichten aus einer EMRFS SQS Amazon-Warteschlange zu lesen, geben Sie den `read-sqs` Befehl ein und geben Sie einen Ausgabespeicherort im lokalen Dateisystem des Master-Knotens für die resultierende Ausgabedatei an.

Sie können eine EMRFS SQS Amazon-Warteschlange auch mit dem `delete-sqs` Befehl löschen.

1. Um Nachrichten aus einer SQS Amazon-Warteschlange zu lesen, geben Sie den folgenden Befehl ein. Ersetzen *queuename* mit dem Namen der SQS Amazon-Warteschlange, die Sie konfiguriert und ersetzt haben */path/filename* mit dem Pfad zur Ausgabedatei:

```
emrfs read-sqs --queue-name queuename --output-file /path/filename
```

Um beispielsweise SQS Amazon-Nachrichten aus der Standardwarteschlange zu lesen und auszugeben, geben Sie Folgendes ein:

```
emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /path/filename
```

Note

Sie können auch die Verknüpfungen `-q` und `-o` anstelle von `--queue-name` und `--output-file` verwenden.

2. Um eine SQS Amazon-Warteschlange zu löschen, geben Sie den folgenden Befehl ein:

```
emrfs delete-sqs --queue-name queuename
```

Um z. B. die Standardwarteschlange zu löschen, geben Sie Folgendes ein:

```
emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU
```

Note

Sie können auch das Kürzel `-q` anstelle von `--queue-name` verwenden.

Konfigurieren der konsistenten Ansicht

Sie können zusätzliche Einstellungen für die konsistente Ansicht konfigurieren, indem Sie sie mithilfe von Konfigurationseigenschaften für `emrfs-site`-Eigenschaften bereitstellen. Sie können beispielsweise einen anderen standardmäßigen DynamoDB-Durchsatz wählen, indem Sie die folgenden Argumente für die CLI `--emrfs` Option angeben, indem Sie die Konfigurationsklassifizierung `emrfs-site` (nur Amazon EMR Release Version 4.x und höher) oder eine Bootstrap-Aktion verwenden, um die Datei `emrfs-site.xml` auf dem Master-Knoten zu konfigurieren:

Example Ändern der Standardwerte zum Lesen und Schreiben von Metadaten bei Cluster-Start

```
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
```

Verwenden Sie alternativ die folgende Konfigurationsdatei und speichern Sie sie lokal oder in Amazon S3:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.consistent.metadata.read.capacity": "600",
      "fs.s3.consistent.metadata.write.capacity": "300"
    }
  }
]
```

Verwenden Sie die Konfiguration, die Sie erstellt haben, mit der folgenden Syntax:

```
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

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Die folgenden Optionen können mithilfe von Konfigurationen oder Argumenten festgelegt werden. AWS CLI --emrfs Informationen über diese Argumente finden Sie in der [AWS CLI -Befehlsreferenz](#).

emrfs-site.xml-Eigenschaften für die konsistente Ansicht

Eigenschaft	Standardwert	Beschreibung
fs.s3.consistent	false	

Eigenschaft	Standardwert	Beschreibung
		Wenn diese Eigenschaft auf true gesetzt ist, wird sie so konfiguriert, EMRFS dass DynamoDB verwendet wird, um Konsistenz zu gewährleisten.
<code>fs.s3.consistent.retryPolicyType</code>	exponential	Diese Eigenschaft gibt die Richtlinie an, die für Wiederholversuche bei Konsistenzproblemen verwendet werden soll. Optionen: exponentiell, fest oder keine.
<code>fs.s3.consistent.retryPeriodSeconds</code>	1	Diese Eigenschaft legt die Wartezeit zwischen den Wiederholversuchen bei Konsistenzproblemen fest.
<code>fs.s3.consistent.retryCount</code>	10	Diese Eigenschaft legt die maximale Anzahl von Wiederholversuchen bei erkannten Inkonsistenzen fest.
<code>fs.s3.consistent.throwExceptionOnInconsistency</code>	true	Diese Eigenschaft bestimmt, ob eine Konsistenzausnahme ausgelöst oder protokolliert werden soll. Wenn diese Eigenschaft auf true festgelegt wurde, wird eine <code>ConsistencyException</code> ausgelöst.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent.metadata.autoCreate</code>	true	Wenn diese Eigenschaft auf true festgelegt wurde, ermöglicht sie die automatische Erstellung von Metadaten tabellen.
<code>fs.s3.consistent.metadata.etagVerificationEnabled</code>	true	Mit Amazon EMR 5.29.0 ist diese Eigenschaft standardmäßig aktiviert. Wenn diese Option aktiviert ist, wird S3 EMRFS verwendet, ETags um zu überprüfen, ob es sich bei den gelesenen Objekten um die neueste verfügbare Version handelt. Diese Funktion ist hilfreich für read-after-update Anwendungsfälle, in denen Dateien auf S3 überschrieben werden, während derselbe Name beibehalten wird. Diese ETag Überprüfungsfunktion funktioniert derzeit nicht mit S3 Select.
<code>fs.s3.consistent.metadata.tableName</code>	EmrFSMetadata	Diese Eigenschaft gibt den Namen der Metadaten tabelle in DynamoDB an.
<code>fs.s3.consistent.metadata.readCapacity</code>	500	Diese Eigenschaft gibt die DynamoDB-Lesekapazität an, die bei Erstellung der Metadaten tabelle bereitgestellt wird.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent.metadata.write.capacity</code>	100	Diese Eigenschaft gibt die DynamoDB-Schreibkapazität an, die bei Erstellung der Metadaten tabelle bereitgestellt wird.
<code>fs.s3.consistent.fastList</code>	true	Wenn diese Eigenschaft auf true festgelegt wurde, verwendet sie mehrere Threads zum Auflisten eines Verzeichnisses (wenn notwendig). Damit diese Eigenschaft verwendet werden kann, muss Konsistenz aktiviert sein.
<code>fs.s3.consistent.fastList.prefetchMetadata</code>	false	Wenn diese Eigenschaft auf true festgelegt wurde, ermöglicht sie den Vorabruf von Metadaten für Verzeichnisse mit mehr als 20.000 Elementen.
<code>fs.s3.consistent.notification.CloudWatch</code>	false	Wenn diese Option auf gesetzt ist true , werden CloudWatch Metriken für FileSystem API Aufrufe aktiviert, die aufgrund eventueller Konsistenzprobleme in Amazon S3 fehlschlagen.
<code>fs.s3.consistent.notification.SQS</code>	false	Wenn diese Option aktiviert ist true , werden eventuelle Konsistenzbenachrichtigungen in eine SQS Amazon-Warteschlange verschoben.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent.notification.SQS.queueName</code>	EMRFS-Inconsistency- <jobFlowId>	Wenn Sie diese Eigenschaft ändern, können Sie Ihren eigenen SQS Warteschlangenamen für Nachrichten angeben, die sich auf etwaige Konsistenzprobleme in Amazon S3 beziehen.
<code>fs.s3.consistent.notification.SQS.customMsg</code>	none	Mit dieser Eigenschaft können Sie benutzerdefinierte Informationen angeben, die in SQS Nachrichten zu eventuellen Konsistenzproblemen in Amazon S3 enthalten sind. Wenn für diese Eigenschaft kein Wert festgelegt ist, ist das entsprechende Feld in der Nachricht leer.
<code>fs.s3.consistent.dynamodb.endpoint</code>	none	Über diese Eigenschaft können Sie einen benutzerdefinierten DynamoDB-Endpunkt für Ihre Metadaten der konsistenten Ansicht angeben.
<code>fs.s3.useRequesterPaysHeader</code>	false	Wenn diese Eigenschaft auf true gesetzt ist, erlaubt sie Amazon-S3-Anfragen an Buckets mit aktivierter Bezahler-Option.

EMRFSCLIBefehlsreferenz

Der EMRFS CLI ist standardmäßig auf allen Cluster-Masterknoten installiert, die mit der EMR Amazon-Version 3.2.1 oder höher erstellt wurden. Sie können den verwenden EMRFSCLI, um die Metadaten für eine konsistente Ansicht zu verwalten.

Note

Der `emrfs` Befehl wird nur mit der VT1 00-Terminalemulation unterstützt. Er funktioniert jedoch unter Umständen auch mit anderen Terminal-Emulator-Modi.

Übergeordneter Befehl "emrfs"

Der Top-Level-Befehl `emrfs` unterstützt die folgende Struktur.

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
 \
 list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

Geben Sie [Optionen] mit oder ohne [Argumente] an, wie in der folgenden Tabelle beschrieben. Für [Optionen], die speziell für Unterbefehle (`describe-metadata`, `set-metadata-capacity` usw.) gelten, finden Sie nachstehend die jeweiligen Unterbefehle.

[Optionen] für `emrfs`

Option	Beschreibung	Erforderlich
<code>-a <i>AWS_ACCESS_KEY_ID</i></code> <code>--access-key <i>AWS_ACCESS_KEY_ID</i></code>	Der AWS Zugriffsschlüssel, den Sie verwenden, um Objekte in Amazon S3 zu schreiben und einen Metadatenpeicher in DynamoDB zu erstellen oder darauf zuzugreifen. Standardmäßig <i>AWS_ACCESS_KEY_ID</i> ist auf den Zugriffsschlüssel festgelegt, der zur Erstellung des Clusters verwendet wurde.	Nein
		Nein

Option	Beschreibung	Erforderlich
<code>-s <i>AWS_SECRET_ACCESS_KEY</i> --secret-key <i>AWS_SECRET_ACCESS_KEY</i></code>	Der AWS geheime Schlüssel, der dem Zugriffsschlüssel zugeordnet ist, den Sie verwenden, um Objekte in Amazon S3 zu schreiben und einen Metadatenpeicher in DynamoDB zu erstellen oder darauf zuzugreifen. Standardmäßig <i>AWS_SECRET_ACCESS_KEY</i> ist auf den geheimen Schlüssel gesetzt, der dem Zugriffsschlüssel zugeordnet ist, der zur Erstellung des Clusters verwendet wurde.	
<code>-v --verbose</code>	Erzeugt eine ausführliche Ausgabe.	Nein
<code>-h --help</code>	Zeigt die Hilfenachricht für den <code>emrfs</code> -Befehl mit einer Verwendungsanweisung an.	Nein

Unterbefehl "emrfs describe-metadata"

[Optionen] für `emrfs describe-metadata`

Option	Beschreibung	Erforderlich
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-Tabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. <code>EmrFSMetadata</code>	Nein

Example Beispiel für "emrfs describe-metadata"

Das folgende Beispiel beschreibt die Standardmetadaten-Tabelle.

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
```

```
write-capacity: 100
status: ACTIVE
approximate-item-count (6 hour delay): 12
```

Unterbefehl `emrfs set-metadata-capacity`

[Optionen] für `emrfs set-metadata-capacity`

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. <code>EmrFSMetadata</code>	Nein
-r <i>READ_CAPACITY</i> --read-capacity <i>READ_CAPACITY</i>	Die angeforderte Lesedurchsatzkapazität für die Metadatentabelle. Wenn das Symbol <i>READ_CAPACITY</i> Argument wurde nicht angegeben, der Standardwert ist <code>400</code> .	Nein
-w <i>WRITE_CAPACITY</i> --write-capacity <i>WRITE_CAPACITY</i>	Die angeforderte Schreibdurchsatzkapazität für die Metadatentabelle. Wenn das Symbol <i>WRITE_CAPACITY</i> Argument wurde nicht angegeben, der Standardwert ist <code>100</code> .	Nein

Example Beispiel für `emrfs set-metadata-capacity`

Im folgenden Beispiel ist für die Metadatentabelle `600` die Lesedurchsatzkapazität auf `150` und die Schreibkapazität auf `EmrMetadataAlt` festgelegt.

```
$ 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
```

Unterbefehl "emrfs delete-metadata"

[Optionen] für emrfs delete-metadata

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. EmrFSMetadata	Nein

Example Beispiel für "emrfs delete-metadata"

Das folgende Beispiel löscht die Standardmetadatentabelle.

```
$ emrfs delete-metadata
```

Unterbefehl "emrfs create-metadata"

[Optionen] für emrfs create-metadata

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. EmrFSMetadata	Nein
-r <i>READ_CAPACITY</i> --read-capacity <i>READ_CAPACITY</i>	Die angeforderte Lesedurchsatzkapazität für die Metadatentabelle. Wenn das Symbol <i>READ_CAPACITY</i> Argument wurde nicht angegeben, der Standardwert ist 400.	Nein
		Nein

Option	Beschreibung	Erforderlich
-w <i>WRITE_CAPACITY</i> --write-capacity <i>WRITE_CAPACITY</i>	Die angeforderte Schreibdurchsatzkapazität für die Metadatentabelle. Wenn das Symbol <i>WRITE_CAPACITY</i> Argument wurde nicht angegeben, der Standardwert ist 100.	

Example Beispiel für "emrfs create-metadata"

Im folgenden Beispiel wird eine Metadatentabelle mit dem Namen EmrFSMetadataAlt erstellt.

```
$ emrfs create-metadata -m EmrFSMetadataAlt
Creating metadata: EmrFSMetadataAlt
EmrFSMetadataAlt
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 0
```

Unterbefehl emrfs list-metadata-stores

Der Unterbefehl emrfs list-metadata-stores hat keine [Optionen].

Example L ist-metadata-stores Beispiel

Im folgenden Beispiel werden Ihre Metadatentabellen aufgelistet.

```
$ emrfs list-metadata-stores
EmrFSMetadata
```

Unterbefehl "emrfs diff"

[Optionen] für emrfs diff

Option	Beschreibung	Erforderlich
	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_</i>	Nein

Option	Beschreibung	Erforderlich
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>NAME</i> Argument wurde nicht angegeben, der Standardwert ist. <code>EmrFSMetadata</code>	
<code>s3://s3Path</code>	Der Pfad zu dem Amazon-S3-Bucket für den Vergleich mit der Metadaten-Tabelle. Die Buckets werden rekursiv synchronisiert.	Ja

Example Beispiel für "emrfs diff"

Im folgenden Beispiel wird die Standardmetadaten-Tabelle mit einem Amazon-S3-Bucket verglichen.

```
$ 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
```

Unterbefehl "emrfs delete"

[Optionen] für emrfs delete

Option	Beschreibung	Erforderlich
		Nein

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. <code>EMRFSMetadata</code>	
<i>s3://s3Path</i>	Der Pfad zum Amazon-S3-Bucket, den Sie für die konsistente Ansicht verfolgen. Die Buckets werden rekursiv synchronisiert.	Ja
-t <i>TIME</i> --time <i>TIME</i>	Die Ablaufzeit (interpretiert mithilfe des Zeiteinheit-Arguments). Alle Metadateneinträge, die älter sind als <i>TIME</i> Das Argument wird für den angegebenen Bucket gelöscht.	
-u <i>UNIT</i> --time-unit <i>UNIT</i>	Die Maßeinheit für die Interpretation des Zeitarguments (Nanosekunden, Mikrosekunden, Millisekunden, Sekunden, Minuten, Stunden oder Tage). Wenn kein Argument angegeben ist, lautet der Standardwert <code>days</code> .	
--read-consumption <i>READ_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Lesedurchsatz für die Operation <code>delete</code> . Wenn das Symbol <i>READ_CONSUMPTION</i> Argument ist nicht angegeben, der Standardwert ist <code>400</code> .	Nein
--write-consumption <i>WRITE_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Schreibdurchsatz für die Operation <code>delete</code> . Wenn das Symbol <i>WRITE_CONSUMPTION</i> Argument ist nicht angegeben, der Standardwert ist <code>100</code> .	Nein

Example Beispiel für "emrfs delete"

Im folgenden Beispiel werden alle Objekte in einem Amazon-S3-Bucket aus den verfolgenden Metadaten für die konsistente Ansicht entfernt.

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

Unterbefehl "emrfs import"

[Optionen] für emrfs import

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. <code>EmrFSMetadata</code>	Nein
<i>s3://s3Path</i>	Der Pfad zum Amazon-S3-Bucket, den Sie für die konsistente Ansicht verfolgen. Die Buckets werden rekursiv synchronisiert.	Ja
--read-consumption <i>READ_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Lesedurchsatz für die Operation delete. Wenn das Symbol <i>READ_CONSUMPTION</i> Argument ist nicht angegeben, der Standardwert ist <code>400</code> .	Nein
--write-consumption <i>WRITE_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Schreibdurchsatz für die Operation delete. Wenn das Symbol <i>WRITE_CONSUMPTION</i> Argument ist nicht angegeben, der Standardwert ist <code>100</code> .	Nein

Example Beispiel für "emrfs import"

Im folgenden Beispiel werden alle Objekte in einem Amazon-S3-Bucket mit den verfolgenden Metadaten für die konsistente Ansicht importiert. Alle unbekannt Schlüssel werden ignoriert.

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

Unterbefehl "emrfs sync"

[Optionen] für emrfs sync

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Symbol <i>METADATA_NAME</i> Argument wurde nicht angegeben, der Standardwert ist. <code>EmrFSMetadata</code>	Nein
<i>s3://s3Path</i>	Der Pfad zum Amazon-S3-Bucket, den Sie für die konsistente Ansicht verfolgen. Die Buckets werden rekursiv synchronisiert.	Ja
--read-consumption <i>READ_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Lesedurchsatz für die Operation delete. Wenn das Symbol <i>READ_CONSUMPTION</i> Argument ist nicht angegeben, der Standardwert ist <code>400</code> .	Nein
--write-consumption <i>WRITE_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Schreibdurchsatz für die Operation delete. Wenn das Symbol <i>WRITE_CONSUMPTION</i> Argument ist nicht angegeben, der Standardwert ist <code>100</code> .	Nein

Example Beispiel für "emrfs sync"

Im folgenden Beispiel werden alle Objekte in einem Amazon-S3-Bucket mit den verfolgenden Metadaten für die konsistente Ansicht importiert. Alle unbekannt Schlüssel werden gelöscht.

```
$ 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
```

Unterbefehl "emrfs read-sqs"

[Optionen] für emrfs read-sqs

Option	Beschreibung	Erforderlich
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> ist der Name der SQS Amazon-Warteschlange, die in konfiguriert ist <code>emrfs-site.xml</code> . Der Standardwert ist EMRFS-Inc consistency-<jobFlowId> .	Ja
<code>-o <i>OUTPUT_FILE</i> --output-file <i>OUTPUT_FILE</i></code>	<i>OUTPUT_FILE</i> ist der Pfad zur Ausgabedatei im lokalen Dateisystem des Master-Knotens. Nachrichten, die aus der Warteschlange gelesen werden, werden in diese Datei geschrieben.	Ja

Unterbefehl "emrfs delete-sqs"

[Optionen] für emrfs delete-sqs

Option	Beschreibung	Erforderlich
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> ist der Name der SQS Amazon-Warteschlange, die in <code>emrfs-site.xml</code> konfiguriert ist. Der Standardwert ist EMRFS-Inconsistency-<jobFlowId> .	Ja

EMRFSCLIBefehle als Schritte einreichen

Das folgende Beispiel zeigt, wie Sie das `emrfs` Hilfsprogramm auf dem Master-Knoten verwenden, indem Sie die Befehle AWS CLI oder API und die `verwendencommand-runner.jar`, um den `emrfs` Befehl als Schritt auszuführen. Das Beispiel verwendet den AWS SDK for Python (Boto3), um einem Cluster einen Schritt hinzuzufügen, der Objekte in einem Amazon S3 S3-Bucket zur EMRFS Standard-Metadatentabelle hinzufügt.

```
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()
```

Sie können den zurückgegebenen Wert `step_id` verwenden, um die Protokolle auf das Ergebnis der Operation zu prüfen.

Autorisieren des Zugriffs auf EMRFS Daten in Amazon S3

Standardmäßig EC2 bestimmt die EMR Rolle für die Berechtigungen für den Zugriff auf EMRFS Daten in Amazon S3. Die mit dieser Rolle verknüpften IAM Richtlinien gelten unabhängig davon, über welchen Benutzer oder welche Gruppe die Anfrage gestellt hat EMRFS. Der Standardwert ist `EMR_EC2_DefaultRole`. Weitere Informationen finden Sie unter [Servicerolle für EC2 Clusterinstanzen \(EC2Instanzprofil\)](#).

Ab der EMR Amazon-Release-Version 5.10.0 können Sie eine Sicherheitskonfiguration verwenden, um IAM Rollen für EMRFS festzulegen. Auf diese Weise können Sie die Berechtigungen für EMRFS Anfragen an Amazon S3 für Cluster mit mehreren Benutzern anpassen. Sie können verschiedene IAM Rollen für verschiedene Benutzer und Gruppen sowie für verschiedene Amazon S3-Bucket-Standorte basierend auf dem Präfix in Amazon S3 angeben. Wenn EMRFS eine Anfrage an Amazon S3 gestellt wird, die den von Ihnen angegebenen Benutzern, Gruppen oder Standorten entspricht, verwendet der Cluster die entsprechende Rolle, die Sie angeben, anstelle der EMR Rolle für EC2. Weitere Informationen finden [Sie unter IAM Rollen für EMRFS Anfragen an Amazon S3 konfigurieren](#).

Wenn Ihre EMR Amazon-Lösung Anforderungen EMRFS stellt, die über die IAM Rollen hinausgehen, können Sie alternativ eine Anbieterklasse für benutzerdefinierte Anmeldeinformationen definieren, mit der Sie den Zugriff auf EMRFS Daten in Amazon S3 anpassen können.

Einen Anbieter für benutzerdefinierte Anmeldeinformationen für EMRFS Daten in Amazon S3 erstellen

Um einen Anbieter für benutzerdefinierte Anmeldeinformationen zu erstellen, implementieren Sie die Klassen [AWSCredentialsProvider](#) und Hadoop [Configurable](#).

Eine ausführliche Erläuterung dieses Ansatzes finden Sie unter [Daten aus einem anderen AWS Konto sicher analysieren mit EMRFS](#) im AWS Big Data-Blog. Der Blogbeitrag enthält ein Tutorial, das Sie durch den Prozess führt end-to-end, von der Erstellung von IAM Rollen bis hin zum Starten des Clusters. Er enthält auch ein Java-Codebeispiel zum Implementieren der benutzerdefinierten Anmeldeinformationsanbieter-Klasse.

Die grundlegenden Schritte sind wie folgt:

So definieren Sie einen benutzerdefinierten Anmeldeinformationsanbieter

1. Erstellen Sie eine benutzerdefinierte Anbieterklasse für Anmeldeinformationen, die als JAR Datei kompiliert wurde.
2. Führen Sie ein Skript als Bootstrap-Aktion aus, um die JAR Anbieterdatei für benutzerdefinierte Anmeldeinformationen an den `/usr/share/aws/emr/emrfs/auxlib` Speicherort auf dem Master-Knoten des Clusters zu kopieren. Weitere Informationen zu Bootstrap-Aktionen finden Sie unter [\(Optional\) Erstellen von Bootstrap-Aktionen zum Installieren zusätzlicher Software](#).
3. Passen Sie die `emrfs-site` Klassifizierung an, um die Klasse anzugeben, die Sie in der JAR Datei implementieren. Weitere Informationen zur Angabe von Konfigurationsobjekten

zur Anpassung von Anwendungen finden Sie unter [Konfiguration von Anwendungen](#) im EMR Amazon-Versionshandbuch.

Das folgende Beispiel zeigt einen `create-cluster`-Befehl, der einen Hive-Cluster mit üblichen Konfigurationsparametern startet und außerdem Folgendes umfasst:

- Eine Bootstrap-Aktion, die das Skript `copy_jar_file.sh` ausführt, das in `mybucket` in Amazon S3 gespeichert wird.
- Eine `emrfs-site` Klassifizierung, die einen Anbieter für benutzerdefinierte Anmeldeinformationen spezifiziert, der in der JAR Datei wie folgt definiert ist `MyCustomCredentialsProvider`

Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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":[]}]'
```

Den AWS Security Token Service Standardendpunkt verwalten

EMRFS verwendet die AWS Security Token Service (STS), um temporäre Sicherheitsanmeldedaten für den Zugriff auf Ihre AWS Ressourcen abzurufen. Frühere EMR Amazon-Release-Versionen

senden alle AWS STS Anfragen an einen einzigen globalen Endpunkt unter `https://sts.amazonaws.com`. Amazon EMR veröffentlicht die Versionen 5.31.0 und 6.1.0 und höher und sendet stattdessen Anfragen an regionale AWS STS Endpunkte. Dies reduziert die Latenz und verbessert die Gültigkeit von Sitzungstoken. Weitere Informationen zu AWS STS Endpunkten finden Sie im Benutzerhandbuch unter [Verwaltung AWS STS in einer AWS Region](#). AWS Identity and Access Management

Wenn Sie die EMR Amazon-Release-Versionen 5.31.0 und 6.1.0 und höher verwenden, können Sie den Standardendpunkt überschreiben. AWS STS Dazu müssen Sie die `fs.s3.sts.endpoint`-Eigenschaft in Ihrer `emrfs-site`-Konfiguration ändern.

Im folgenden AWS CLI Beispiel wird der von verwendete AWS STS Standardendpunkt EMRFS auf den globalen Endpunkt gesetzt.

```
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

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

Alternativ können Sie mit dem folgenden Beispiel eine JSON Konfigurationsdatei erstellen und diese mit dem `--configurations` Argument von `aws emr create-cluster` angeben. Weitere Informationen zur Verwendung `--configurations`, finden Sie unter [AWS CLI in der Command-Befehlsreferenz](#).

```
[  
  {  
    "classification": "emrfs-site",  
    "properties": {  
      "fs.s3.sts.endpoint": "https://sts.amazonaws.com"  
    }  
  }  
]
```

Amazon S3 S3-Verschlüsselung mithilfe von EMRFS Eigenschaften angeben

Important

Ab der EMR Amazon-Release-Version 4.8.0 können Sie Sicherheitskonfigurationen verwenden, um Verschlüsselungseinstellungen einfacher und mit mehr Optionen anzuwenden. Wir empfehlen die Verwendung von Sicherheitskonfigurationen. Weitere Informationen finden Sie unter [Datenverschlüsselung konfigurieren](#). Die in diesem Abschnitt beschriebenen Konsolenbefehle sind für Versionen vor 4.8.0 verfügbar. Wenn Sie die AWS CLI Amazon S3 S3-Verschlüsselung sowohl in der Cluster-Konfiguration als auch in einer Sicherheitskonfiguration in nachfolgenden Versionen verwenden, überschreibt die Sicherheitskonfiguration die Cluster-Konfiguration.

Wenn Sie einen Cluster erstellen, können Sie serverseitige Verschlüsselung (SSE) oder clientseitige Verschlüsselung (CSE) für EMRFS Daten in Amazon S3 mithilfe der Konsole oder mithilfe von `emrfs-site` Klassifizierungseigenschaften über das oder angeben. AWS CLI EMR SDK Amazon S3 SSE und schließen CSE sich gegenseitig aus. Sie können eines wählen, aber nicht beide.

AWS CLI Anweisungen finden Sie unten im entsprechenden Abschnitt für Ihren Verschlüsselungstyp.

Um EMRFS Verschlüsselungsoptionen anzugeben, verwenden Sie AWS Management Console

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie Release (Version) 4.7.2 oder früher aus.
4. Wählen Sie weitere Optionen für Software and Steps (Software und Steps) aus wie für Ihre Anwendung erforderlich. Wählen Sie anschließend Next (Weiter) aus.
5. Wählen Sie in den Bereichen Hardware (Hardware) und General Cluster Settings (Allgemeine Cluster-Einstellungen) Einstellungen aus wie für Ihre Anwendung erforderlich.
6. Wählen Sie im Bereich Sicherheit unter Authentifizierung und Verschlüsselung die zu verwendende Option S3-Verschlüsselung (mitEMRFS) aus.

Note

Die serverseitige S3-Verschlüsselung mit KMS Key Management (SSE-KMS) ist nicht verfügbar, wenn Sie die EMR Amazon-Version 4.4 oder früher verwenden.

- Wenn Sie eine Option wählen, die AWS Key Management verwendet, wählen Sie eine AWS KMS Schlüssel-ID. Weitere Informationen finden Sie unter [Wird für die Verschlüsselung verwendet AWS KMS keys EMRFS](#).
 - Wenn Sie die clientseitige S3-Verschlüsselung mit benutzerdefiniertem Materialanbieter wählen, geben Sie den Klassennamen und den JAR Speicherort an. Weitere Informationen finden Sie unter [Clientseitige Verschlüsselung für Amazon S3](#).
7. Wählen Sie weitere Optionen wie für Ihre Anwendung erforderlich aus. Wählen Sie anschließend Create Cluster (Cluster erstellen) aus.

Wird für die Verschlüsselung verwendet AWS KMS keys EMRFS

Der AWS KMS Verschlüsselungsschlüssel muss in derselben Region wie Ihre EMR Amazon-Cluster-Instance und die verwendeten Amazon S3-Buckets erstellt werden. EMRFS Wenn sich der von Ihnen angegebene Schlüssel in einem anderen Konto befindet als dem, das Sie zur Konfiguration eines Clusters verwenden, müssen Sie den Schlüssel mit seinem ARN angeben.

Die Rolle für das EC2 Amazon-Instance-Profil muss über Berechtigungen zur Verwendung des von Ihnen angegebenen KMS Schlüssels verfügen. Die Standardrolle für das Instance-Profil in Amazon EMR ist `EMR_EC2_DefaultRole`. Wenn Sie eine andere Rolle für das Instance-Profil oder IAM Rollen für EMRFS Anfragen an Amazon S3 verwenden, stellen Sie sicher, dass jede Rolle je nach Bedarf als Schlüsselbenutzer hinzugefügt wird. Dadurch erhält die Rolle die Erlaubnis, den KMS Schlüssel zu verwenden. Weitere Informationen finden Sie unter [Using Key Policies](#) im AWS Key Management Service Developer Guide und [Configure IAM roles for EMRFS requests to Amazon S3](#).

Sie können das verwenden AWS Management Console , um Ihr Instance-Profil oder EC2 Instance-Profil zur Liste der Schlüsselbenutzer für den angegebenen KMS Schlüssel hinzuzufügen, oder Sie können das AWS CLI oder an verwenden, AWS SDK um eine entsprechende Schlüsselrichtlinie anzuhängen.

Beachten Sie, dass Amazon nur [symmetrische KMS Schlüssel EMR](#) unterstützt. Sie können keinen [asymmetrischen KMS Schlüssel](#) verwenden, um ruhende Daten in einem EMR Amazon-Cluster zu

verschlüsseln. Hilfe bei der Bestimmung, ob ein KMS Schlüssel symmetrisch oder asymmetrisch ist, finden Sie unter [Identifizieren symmetrischer](#) und asymmetrischer Schlüssel. KMS

Im Folgenden wird beschrieben, wie Sie das standardmäßige EMR Amazon-Instance-Profil `EMR_EC2_DefaultRole` als Hauptbenutzer mithilfe von hinzufügen AWS Management Console. Es wird davon ausgegangen, dass Sie bereits einen KMS Schlüssel erstellt haben. Informationen zum Erstellen eines neuen KMS Schlüssels finden Sie unter [Schlüssel erstellen](#) im AWS Key Management Service Entwicklerhandbuch.

Um das EC2 Instance-Profil für Amazon EMR zur Liste der Benutzer von Verschlüsselungsschlüsseln hinzuzufügen

1. Melden Sie sich bei der AWS Key Management Service (AWS KMS) -Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/kms>.
2. Um das zu ändern AWS-Region, verwenden Sie die Regionsauswahl in der oberen rechten Ecke der Seite.
3. Wählen Sie den Alias des KMS Schlüssels aus, den Sie ändern möchten.
4. Wählen Sie auf der Seite mit den Schlüsseldetails unter Key Users (Schlüsselbenutzer(die Option Add (Hinzufügen) aus.
5. Wählen Sie die entsprechende Rolle im Dialogfeld Add key users (Schlüsselbenutzer hinzufügen) aus. Der Name der Standardrolle lautet `EMR_EC2_DefaultRole`.
6. Wählen Sie Hinzufügen aus.

Serverseitige Verschlüsselung im Amazon S3

Wenn Sie die Amazon-S3-Verschlüsselung einrichten, verschlüsselt Amazon S3 die Daten auf der Objektebene, während die Daten auf den Datenträger geschrieben werden, und entschlüsselt sie, wenn auf sie zugegriffen wird. Weitere Informationen SSE dazu finden Sie unter [Schützen von Daten mithilfe serverseitiger Verschlüsselung](#) im Amazon Simple Storage Service-Benutzerhandbuch.

Bei der Angabe SSE in Amazon können Sie zwischen zwei verschiedenen Schlüsselverwaltungssystemen wählenEMR:

- SSE-S3 — Amazon S3 verwaltet Schlüssel für Sie.
- SSE- KMS — Sie verwenden eine AWS KMS key , um Richtlinien einzurichten, die für Amazon geeignet sindEMR. Weitere Informationen zu den wichtigsten Anforderungen für Amazon EMR finden Sie unter [AWS KMS keys Zur Verschlüsselung verwenden](#).

SSE mit vom Kunden bereitgestellten Schlüsseln (SSE-C) ist nicht für die Verwendung mit Amazon verfügbar. EMR

Um einen Cluster mit aktiviertem SSE -S3 zu erstellen, verwenden Sie AWS CLI

- Geben Sie den folgenden Befehl ein:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

Sie können SSE -S3 auch aktivieren, indem Sie `fs.s3` setzen. `enableServerSideEncryption` Die Eigenschaft `Encryption` ist in den Eigenschaften auf `true` gesetzt. `emrfs-site` Sehen Sie sich das Beispiel für SSE - KMS unten an und lassen Sie die Eigenschaft für die Schlüssel-ID weg.

Um einen Cluster mit KMS aktiviertem SSE - zu erstellen, verwenden Sie den AWS CLI

Note

SSE- KMS ist nur in der EMR Amazon-Release-Version 4.5.0 und höher verfügbar.

- Geben Sie den folgenden AWS CLI Befehl ein, um einen Cluster mit SSE - zu erstellen KMS, wobei *keyID* ist ein AWS KMS key, zum Beispiel *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]
```

--ODER--

Geben Sie den folgenden AWS CLI Befehl unter Verwendung der `emrfs-site` Klassifizierung ein und stellen Sie eine JSON Konfigurationsdatei mit dem Inhalt bereit, `myConfig.json` der dem folgenden Beispiel ähnelt:

```
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
```

Beispielinhalt myConfigvon.json:

```
[
  {
    "Classification":"emrfs-site",
    "Properties": {
      "fs.s3.enableServerSideEncryption": "true",
      "fs.s3.serverSideEncryption.kms.keyId":"a4567b8-9900-12ab-1234-123a45678901"
    }
  }
]
```

Konfigurationseigenschaften für SSE -S3 und - SSE KMS

Diese Eigenschaften können unter Verwendung der `emrfs-site`- Konfigurationsklassifikation konfiguriert werden. SSE- KMS ist nur in der EMR Amazon-Release-Version 4.5.0 und höher verfügbar.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.enableServerSideEncryption</code>	false	Wenn dies auf true festgelegt wurde, werden in Amazon S3 gespeicherte Objekte mittels serverseitiger Verschlüsselung verschlüsselt. Wenn kein Schlüssel angegeben ist, wird SSE -S3 verwendet.
<code>fs.s3.serverSideEncryption.kms.keyId</code>	n/a	Spezifiziert eine AWS KMS Schlüssel-ID oderARN. Wenn ein Schlüssel angegeben ist, KMS wird SSE - verwendet.

Clientseitige Verschlüsselung für Amazon S3

Bei der clientseitigen Amazon S3 S3-Verschlüsselung erfolgt die Amazon S3 S3-Verschlüsselung und -Entschlüsselung im EMRFS Client auf Ihrem Cluster. Objekte werden vor dem Hochladen nach Amazon S3 verschlüsselt und nach dem Herunterladen entschlüsselt. Der von Ihnen festgelegte Anbieter stellt den vom Client verwendeten Verschlüsselungsschlüssel bereit. Der Client kann die von AWS KMS (CSE-KMS) bereitgestellten Schlüssel oder eine benutzerdefinierte Java-Klasse verwenden, die den clientseitigen Stammschlüssel (-C) bereitstellt. CSE Die Verschlüsselungsspezifikationen zwischen CSE - KMS und CSE -C unterscheiden sich geringfügig, abhängig vom angegebenen Anbieter und den Metadaten des Objekts, das entschlüsselt oder verschlüsselt wird. Weitere Informationen zu diesen Unterschieden finden Sie unter [Schützen von Daten durch clientseitige Verschlüsselung](#) im Entwicklerhandbuch von Amazon Simple Storage Service.

Note

Amazon S3 stellt CSE lediglich sicher, dass die mit Amazon S3 ausgetauschten EMRFS Daten verschlüsselt sind. Nicht alle Daten auf Cluster-Instance-Volumes sind verschlüsselt. Da Hue es nicht verwendet EMRFS, werden Objekte, die der Hue S3 File Browser in Amazon S3 schreibt, außerdem nicht verschlüsselt.

Zur CSE Spezifizierung — KMS für EMRFS Daten in Amazon S3 mit dem AWS CLI

- Geben Sie den folgenden Befehl ein und ersetzen Sie *MyKMSKeyID* mit der Schlüssel-ID oder ARN dem zu verwendenden KMS Schlüssel:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier  
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

Erstellen eines benutzerdefinierten Schlüsselanbieters

Abhängig von der Art der Verschlüsselung, die Sie bei der Erstellung eines benutzerdefinierten Schlüsselanbieters verwenden, muss die Anwendung auch unterschiedliche EncryptionMaterialsProvider Schnittstellen implementieren. Beide Schnittstellen sind in der Version 1.11.0 AWS SDK für Java und höher verfügbar.

- Um die Amazon S3 S3-Verschlüsselung zu implementieren, verwenden Sie das Modell [com.amazonaws.services.s3.model.EncryptionMaterialsProvider Schnittstelle](#).
- Verwenden Sie die Datei [com.amazonaws.services.elasticmapreduce.spi.security](#), um die lokale Festplattenverschlüsselung zu implementieren. [EncryptionMaterialsProvider Schnittstelle](#).

Sie können jede Strategie verwenden, um Verschlüsselungsmaterial für die Implementierung bereitzustellen. Sie können sich beispielsweise dafür entscheiden, statisches Verschlüsselungsmaterial bereitzustellen oder es in ein komplexeres Schlüsselverwaltungssystem zu integrieren.

Wenn Sie die Amazon S3 S3-Verschlüsselung verwenden, müssen Sie die Verschlüsselungsalgorithmen AES/GCM/NoPadding für benutzerdefinierte Verschlüsselungsmaterialien verwenden.

Wenn Sie die lokale Festplattenverschlüsselung verwenden, variiert der Verschlüsselungsalgorithmus, der für benutzerdefinierte Verschlüsselungsmaterialien verwendet werden soll, je nach EMR Version. Für Amazon EMR 7.0.0 und niedriger müssen Sie AES/GCM/NoPadding verwenden. Für Amazon EMR 7.1.0 und höher müssen Sie verwenden AES.

Die EncryptionMaterialsProvider Klasse ruft Verschlüsselungsmaterial nach Verschlüsselungskontext ab. Amazon EMR füllt zur Laufzeit Informationen zum Verschlüsselungskontext aus, damit der Anrufer die richtigen Verschlüsselungsmaterialien für die Rücksendung ermitteln kann.

Example Beispiel: Verwendung eines benutzerdefinierten Schlüsselanbieters für die Amazon S3 S3-Verschlüsselung mit EMRFS

Wenn Amazon die Verschlüsselungsmaterialien von der EncryptionMaterialsProvider Klasse EMR abrufen, um die Verschlüsselung durchzuführen, füllt das materialsDescription Argument EMRFS optional mit zwei Feldern auf: dem Amazon S3 URI für das Objekt und dem des Clusters, die JobFlowId von der EncryptionMaterialsProvider Klasse verwendet werden können, um Verschlüsselungsmaterialien selektiv zurückzugeben.

Beispielsweise kann der Anbieter unterschiedliche Schlüssel für verschiedene Amazon S3 URI S3-Präfixe zurückgeben. Es ist die Beschreibung der zurückgegebenen Verschlüsselungsmaterialien, die letztendlich mit dem Amazon S3 S3-Objekt gespeichert werden, und nicht der materialsDescription Wert, der vom Anbieter generiert EMRFS und an diesen weitergegeben wird. Beim Entschlüsseln eines Amazon S3 S3-Objekts wird die Beschreibung des Verschlüsselungsmaterials an die EncryptionMaterialsProvider Klasse übergeben, sodass sie wiederum selektiv den passenden Schlüssel zur Entschlüsselung des Objekts zurückgeben kann.

Eine `EncryptionMaterialsProvider` Referenzimplementierung finden Sie weiter unten. Ein weiterer benutzerdefinierter Anbieter [EMRFSRSAEncryptionMaterialsProvider](#), ist erhältlich bei 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;
}
}

```

Angeben eines Anbieters für benutzerdefinierte Materialien mithilfe des AWS CLI

Um die AWS CLI zu verwenden, übergeben Sie die Argumente `Encryption`, `ProviderType`, `CustomProviderClass` und `CustomProviderLocation` an die Option `emrfs`.

```

aws emr create-cluster --instance-type m5.xlarge --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=Custom,CustomProviderLocation=s3://
mybucket/myfolder/provider.jar,CustomProviderClass=classname

```

Die Einstellung `ClientSide` aktiviert `Encryption` die clientseitige Verschlüsselung, `CustomProviderClass` ist der Name Ihres `EncryptionMaterialsProvider` Objekts und `CustomProviderLocation` ist der lokale oder Amazon S3 S3-Speicherort, von dem Amazon `CustomProviderClass` auf jeden Knoten im Cluster EMR kopiert und ihn im Klassenpfad platziert.

Angabe eines Anbieters für kundenspezifische Materialien mithilfe eines SDK

Um eine zu verwenden SDK, können Sie die Eigenschaft `fs.s3.cse.encryptionMaterialsProvider.uri` so einrichten, dass die benutzerdefinierte `EncryptionMaterialsProvider` Klasse, die Sie in Amazon S3 speichern, auf jeden Knoten in Ihrem Cluster heruntergeladen wird. Sie konfigurieren dies in einer `emrfs-site.xml` Datei zusammen mit der CSE Option aktiviert und dem richtigen Speicherort des benutzerdefinierten Anbieters.

Bei der AWS SDK for Java Verwendung `RunJobFlowRequest` könnte Ihr Code beispielsweise wie folgt aussehen:

```

<snip>
Map<String,String> emrfsProperties = new HashMap<String,String>();
    emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider.uri", "s3://mybucket/
MyCustomEncryptionMaterialsProvider.jar");

```

```

    emrfsProperties.put("fs.s3.cse.enabled", "true");
    emrfsProperties.put("fs.s3.consistent", "true");

    emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider", "full.class.name.of.EncryptionMaterialsProvider");

    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>

```

EncryptionMaterialsProvider benutzerdefiniert mit Argumenten

Möglicherweise müssen Sie Argumente direkt dem Anbieter übergeben. Zu diesem Zweck können Sie die `emrfs-site`-Konfigurationsklassifizierung mit benutzerdefinierten Argumenten als Eigenschaften verwenden. Eine Beispielkonfiguration ist nachfolgend gezeigt, die als Datei gespeichert wird, `myConfig.json`:

```

[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "myProvider.arg1": "value1",
      "myProvider.arg2": "value2"
    }
  }
]

```


]

Mit dem `create-cluster` Befehl von können Sie die `--configurations` Option verwenden AWS CLI, um die Datei wie unten gezeigt anzugeben:

```
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
```

Konfiguration der EMRFS S3EC V2-Unterstützung

SDKS3-Java-Versionen (1.11.837 und höher) unterstützen den Verschlüsselungsclient Version 2 (S3EC V2) mit verschiedenen Sicherheitsverbesserungen. Weitere Informationen finden Sie im S3-Blogbeitrag: [Updates für den Amazon S3 Encryption Client](#). Weitere Informationen finden Sie im AWS SDK for Java Developer Guide unter [Migration des Amazon S3 S3-Verschlüsselungsclients](#).

Der Verschlüsselungsclient V1 ist aus SDK Gründen der Abwärtskompatibilität weiterhin in der verfügbar. Standardmäßig EMRFS verwendet S3EC V1 zum Verschlüsseln und Entschlüsseln von S3-Objekten, wenn es aktiviert ist. CSE

Mit S3EC V2 verschlüsselte S3-Objekte können nicht EMRFS auf einem EMR Cluster entschlüsselt werden, dessen Release-Version vor `emr-5.31.0` (`emr-5.30.1` und früher, `emr-6.1.0` und früher) liegt.

Example Konfigurieren Sie für die Verwendung von S3EC V2 EMRFS

Um die Verwendung von S3EC V2 EMRFS zu konfigurieren, fügen Sie die folgende Konfiguration hinzu:

```
{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}
```

emrfs-site.xml-Eigenschaften für die Amazon-S3-clientseitige Verschlüsselung

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.cse.enabled</code>	false	Wenn diese Option auf gesetzt ist true , werden in Amazon S3 gespeicherte EMRFS Objekte mit clientseitiger Verschlüsselung verschlüsselt.
<code>fs.s3.cse.encryptionV2.enabled</code>	false	Wenn auf gesetzt true , EMRFS wird der S3-Verschlüsselung client Version 2 zum Verschlüsseln und Entschlüsseln von Objekten auf S3 verwendet. Verfügbar für EMR Version 5.31.0 und höher.
<code>fs.s3.cse.encryptionMaterialsProvider.uri</code>	N/A	Gilt bei der Verwendung benutzerdefinierter Verschlüsselungsmaterialien. Der Amazon S3URI, wo sich der JAR mit dem <code>EncryptionMaterialsProvider</code> befindet. Wenn Sie dies angebenURI, lädt Amazon das EMR automatisch JAR auf alle Knoten im Cluster herunter.
<code>fs.s3.cse.encryptionMaterialsProvider</code>	N/A	Der <code>EncryptionMaterialsProvider</code> -Klassenpfad, der bei der clientseitigen Verschlüsselung verwendet wird. Wenn Sie CSE - verwendenKMS,

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.cse.materialsDescription.enabled</code>	false	<p>Wenn diese Option auf gesetzt ist <code>true</code>, werden die materials Description verschlüsselten Objekte mit dem Amazon S3 URI für das Objekt und dem JobFlowId gefüllt. Bei der Verwendung benutzerdefinierter Verschlüsselungsmaterialien auf <code>true</code> setzen.</p>
<code>fs.s3.cse.kms.keyId</code>	N/A	<p>Gilt bei Verwendung von CSE -KMS. Der Wert des Alias KeyIdARN, oder des für die Verschlüsselung verwendeten KMS Schlüssels.</p>

Eigenschaft	Standardwert	Beschreibung
fs.s3.cse.cryptoStorageMode	ObjectMetadata	<p>Der Amazon-S3-Speichermodus. Standardmäßig wird die Beschreibung der Verschlüsselungsdaten in den Objektmetadaten gespeichert. Sie können die Beschreibung auch in einer Anweisungsdatei speichern. Gültige Werte sind ObjectMetadata und InstructionFile. Weitere Informationen finden Sie unter Clientseitige Datenverschlüsselung mit dem AWS SDK for Java und Amazon S3.</p>

CloudWatch Amazon-Vertreter

Der CloudWatch Amazon-Agent bei Amazon EMR ist ein Tool, das die EC2 Amazon-Instances in Ihrem EMR Cluster überwachen kann. Sie können die Metriken, die Sie mit dem CloudWatch Agenten sammeln, speichern und anzeigen CloudWatch. Weitere Informationen über den CloudWatch Agenten finden Sie im [CloudWatch Amazon-Benutzerhandbuch](#).

Note

Es fallen zusätzliche Gebühren an, wenn Sie andere verwenden, AWS -Services um CloudWatch Amazon-Agentenmetriken zu veröffentlichen, abzufragen oder einzusehen. Weitere Preisinformationen finden Sie auf den folgenden Seiten.

- [CloudWatch Amazon-Preisgestaltung](#)
- [Preise für Amazon Managed Service für Prometheus](#)
- [Amazon Managed Grafana-Preise](#)

Diese Metriken sind unabhängig von den kostenlosen Metriken, die Amazon unter dem AWS/ElasticMapReduce Namespace in EMR CloudWatch bereitstellt. Weitere Informationen zu den Metriken, die der CloudWatch Agent nicht veröffentlicht, finden Sie unter [Überwachen von EMR Amazon-Metriken mit CloudWatch](#).

Mit Amazon EMR 7.0 und höher können Sie eine benutzerdefinierte Version des CloudWatch Amazon-Agenten auf Ihrem Cluster installieren, um Metriken aus Ihrem EMR Cluster zu sammeln. Der Agent sammelt die Metriken von allen Knoten im Cluster, sammelt sie auf dem primären Knoten und veröffentlicht die Metriken von diesem Knoten in der Cloud.

Sie können den Agenten installieren, wenn Sie einen neuen Cluster in der Konsole erstellen oder wenn Sie den `create-cluster` API verwenden. Weitere Informationen finden Sie unter [Erstellen Sie einen EMR Cluster, der Amazon CloudWatch Agent verwendet](#). Standardmäßig veröffentlicht ein EMR Amazon-Cluster, auf dem die Version 7.x-Serie ausgeführt wird, [34 Metriken auf Systemebene](#) in Intervallen von 60 Sekunden. Sie können den Agenten jedoch so konfigurieren, dass er verschiedene Metriken veröffentlicht. CloudWatch Eine weitere Option besteht darin, Metriken in Amazon Managed Service for Prometheus zu veröffentlichen, und Sie können wählen, welche Metriken ebenfalls veröffentlicht werden sollen. Für verschiedene Anwendungsfälle und

Konfigurationen können Sie den Agenten so konfigurieren, dass er die Metriken in der CloudWatch Konsole, Amazon Managed Grafana oder über den APIs for CloudWatch oder Amazon Managed Service for Prometheus anzeigt und abfragt. Für diese AWS -Services fallen Gebühren an, wenn Sie sie zum Speichern und Abfragen von Metriken verwenden.

Der CloudWatch Agent bei Amazon EMR kann Systemmetriken sowie JMX Metriken für die folgenden Dienste in Ihrem EMR Amazon-Cluster veröffentlichen.

- Hadoop DataNode
- Hadoop NameNode
- Garn NodeManager
- Garn ResourceManager
- HBaseMaster — nur Amazon EMR 7.1+
- HBase RegionServer — Nur Amazon EMR 7.1+
- HBase ThriftServer — Nur Amazon EMR 7.1+

Weitere Informationen zu verfügbaren Metriken und zur Konfiguration des CloudWatch Agenten bei Amazon EMR finden Sie unter [CloudWatch Agent für Amazon konfigurieren EMR](#).

In der folgenden Tabelle sind die Version von AmazonCloudWatchAgent aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. AmazonCloudWatchAgent

Informationen zur Version der Komponenten, mit denen AmazonCloudWatchAgent in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

AmazonCloudWatchAgent Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	AmazonCloudWatchAgent Ausführung	Komponenten, die mit installiert wurden 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-

EMRAmazon-Freigabeetikett	AmazonCloudWatchAgent Ausführung	Komponenten, die mit installiert wurden AmazonCloudWatchAgent
		library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server

Themen

- [Erstellen Sie einen EMR Cluster, der Amazon CloudWatch Agent verwendet](#)
- [Standardmetriken für CloudWatch Agenten bei Amazon EMR](#)
- [CloudWatch Agent für Amazon konfigurieren EMR](#)
- [Überlegungen und Einschränkungen](#)
- [CloudWatch Versionsverlauf des Agenten](#)

Erstellen Sie einen EMR Cluster, der Amazon CloudWatch Agent verwendet

Die Verfahren in diesem Abschnitt beschreiben die Schritte zum Erstellen eines Clusters in Amazon EMR mit Amazon CloudWatch Agent aus dem AWS Management Console und dem AWS CLI.

Themen

- [Erforderliche IAM Berechtigungen für den CloudWatch Agenten](#)
- [Erforderlicher CloudWatch Agenten-Endpunkt](#)
- [Erstellen eines EMR-Clusters](#)

Erforderliche IAM Berechtigungen für den CloudWatch Agenten

Der CloudWatch Agent benötigt die AWS Identity and Access Management (IAM) `cloudwatch:PutMetricData` -Berechtigung im EC2 Amazon-Instanzprofil für AmazonEMR. Die

EMR Amazon-Standardrolle hat diese Berechtigung bereits. Sie können die Standardrolle von AWS CLI with aus erstellen `aws emr create-default-roles`. Weitere Informationen finden Sie unter [Servicerolle für EC2 Cluster-Instances \(EC2Instance-Profil\)](#) im Amazon EMR Management Guide.

Die folgende IAM Beispielrichtlinie beinhaltet die `cloudwatch:PutMetricData` Erlaubnis:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "cloudwatch:PutMetricData",
      "Resource": "*"
    }
  ]
}
```

Erforderlicher CloudWatch Agenten-Endpunkt

Um Metriken CloudWatch für einen EMR Cluster in einem privaten Subnetz zu veröffentlichen, erstellen Sie einen CloudWatch Agenten-Endpunkt und ordnen Sie ihn dem zu, in dem sich VPC das private Subnetz befindet.

Weitere Informationen zu den jeweiligen CloudWatch AWS-Region Endpunkten finden Sie unter [CloudWatch Amazon-Endpunkte und Kontingente](#) im AWS Allgemeinen Referenzhandbuch.

Erstellen eines EMR-Clusters

Nachdem Sie die erforderlichen Berechtigungen und den Endpunkt für die Verwendung mit dem CloudWatch Agenten eingerichtet haben, verwenden Sie den AWS Management Console oder, um einen neuen Cluster AWS CLI zu erstellen, auf dem der Agent installiert ist.

Console

Um einen Cluster mit Amazon CloudWatch Agent von der Konsole aus zu erstellen

1. Navigieren Sie zur EMR Amazon-Konsole.
2. Wählen Sie Cluster erstellen.
3. Wählen Sie unter Name und Anwendungen eine EMR Amazon-Version von 7.0.0 oder höher aus.

4. Wählen Sie unter Anwendungspaket das Paket oder die Apps aus, die Sie in Ihrem Cluster installieren möchten, und schließen Sie CloudWatch Agenten in Ihre Auswahl ein.
5. Fahren Sie mit der Erstellung des Clusters fort, um den Anforderungen Ihres Anwendungsfalls gerecht zu werden.

AWS CLI

In der AWS CLI können Sie einen CloudWatch Amazon-Agenten mit dem `--applications` Parameter für zu einem Cluster hinzufügen `create-cluster`.

Um einen Cluster mit CloudWatch Amazon-Agent aus dem AWS CLI

- Wenn Sie einen Cluster erstellen, verwenden Sie einen Befehl, der dem folgenden ähnelt, um den CloudWatch Amazon-Agenten einzubeziehen. *myKey* Ersetzen Sie es durch den Namen Ihres 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
```

Weitere Informationen zur Verwendung von Amazon EMR mit dem AWS CLI finden Sie in der [AWS CLI Befehlsreferenz](#).

Standardmetriken für CloudWatch Agenten bei Amazon EMR

Wenn Sie den CloudWatch Amazon-Agenten auf Amazon installieren EMR, veröffentlicht die Standardkonfiguration die folgenden Systemmetriken für alle Instances in Ihrem Cluster, sofern Sie [den Agenten nicht anders konfigurieren](#). Definitionen der einzelnen [Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken](#) im CloudWatch Amazon-Benutzerhandbuch.

CPU

CPU Metriken

- `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

Festplattenmetriken

- `disk_free`
- `disk_total`
- `disk_used`
- `disk_used_percent`

Memory

Speichermetriken

- `mem_active`
- `mem_available`
- `mem_available_percent`
- `mem_free`
- `mem_inactive`
- `mem_total`
- `mem_used`
- `mem_used_percent`
- `mem_buffered`
- `mem_cached`

Network IO

Netzwerk-I/O-Metriken

- `net_bytes_recv`
- `net_bytes_sent`
- `net_packets_recv`
- `net_packets_sent`

Process

Metriken verarbeiten

- `processes_running`
- `processes_total`

Swap

Metriken wechseln

- `swap_free`
- `swap_used`
- `swap_used_percent`

Standardmäßig veröffentlicht der Agent alle Metriken CloudWatch unter dem benutzerdefinierten Namespace `CWAgent` und unter dem Schema `cluster.id,,instance.id,node.type,service.name`. Gehen Sie wie folgt vor, um diese Standardmetriken anzuzeigen, die der CloudWatch Agent für Amazon veröffentlichtEMR:

1. Navigieren Sie zur CloudWatch Amazon-Konsole.
2. Wählen Sie `Metrics` (Metriken) und dann `All metrics` (Alle Metriken) aus.
3. Wählen Sie unter `Custom Namespaces` (Benutzerdefinierte Namespaces) **CWAgent** aus. Wählen Sie dann das Schema aus: **cluster.id,instance.id,node.type,service.name**.
4. Fragen Sie die Metriken weiterhin CloudWatch über die Benutzeroberfläche ab, wie es Ihr Anwendungsfall erfordert.

Amazon EMR fügt jeder Metrik die folgenden Bezeichnungen hinzu: ["cluster.id", "instance.id", "node.type", "service.name"]. Beachten Sie in Bezug auf diese Labels Folgendes:

- Der Wert des `cluster.id` Labels ist die ID des EMR Clusters, der die Metrik erstellt hat. Ein Beispielwert für das `cluster.id`-Label ist: `j-123456789ABC`.
- Der Wert des `instance.id` Labels ist die ID der Instanz im EMR Cluster, die die Metrik erzeugt hat. Ein Beispielwert für das `instance.id`-Label ist `i-01bcf5f140f3355777`.
- Der Wert des `node.type` Labels steht für den Knotentyp der Instanz im Cluster, die die Metrik erzeugt hat. Dieser Wert kann entweder `mastercore`, oder `seintask`.
- Der Wert des `service.name`-Labels ist der Name des Services, der die Metrik erzeugt hat. Der `service.name`-Standardwert für die Standardsystemmetriken ist `system`.

Note

EMRAmazon-Version 7.0.0 verwendet ein etwas anderes Metrik-Label-Schema. `cluster.id` ist `jobflow.id` und `node.type` ist nicht verfügbar. `instance.id` und `service.name` sind verfügbar.

CloudWatch Agent für Amazon konfigurieren EMR

Amazon EMR 7.0.0 und höher beinhalten den CloudWatch Amazon-Agenten. In den folgenden Abschnitten wird beschrieben, wie Sie den Agenten auf Ihrem EMR Cluster konfigurieren, wie unter [beschrieben Erstellen Sie einen EMR Cluster, der Amazon CloudWatch Agent verwendet](#).

Themen

- [CloudWatch Agent für Amazon EMR 7.1.0 konfigurieren](#)
- [CloudWatch Agent für Amazon EMR 7.0.0 konfigurieren](#)

CloudWatch Agent für Amazon EMR 7.1.0 konfigurieren

Ab Amazon EMR 7.1.0 können Sie den CloudWatch Amazon-Agenten mithilfe der Amazon-Konfiguration so konfigurieren, dass er zusätzliche Systemmetriken verwendet, Anwendungsmetriken hinzufügt und das Metrikziel ändert. EMR API Weitere Informationen darüber, wie Sie die EMR

Konfiguration API zur Konfiguration der Anwendungen Ihres Clusters verwenden können, finden [Sie unter Anwendungen konfigurieren](#).

Note

7.1.0 unterstützt nur den Rekonfigurationstyp. OVERWRITE Weitere Informationen zu den Rekonfigurationstypen finden [Sie unter Überlegungen bei der Neukonfiguration](#) einer Instanzgruppe.

Themen

- [Konfigurationsschema](#)
- [Beispiele für Konfigurationen von Systemmetriken](#)
- [Beispiele für Konfigurationen von Anwendungsmetriken](#)
- [Beispiel für Amazon Managed Service für Prometheus](#)

Konfigurationsschema

`emr-metricshat` die folgenden Klassifizierungen:

- `emr-system-metrics`— konfiguriert Systemmetriken wie CPU Festplatte und Arbeitsspeicher.
- `emr-hadoop-hdfs-datanode-metrics`— Hadoop-Metriken DataNode JMX konfigurieren
- `emr-hadoop-hdfs-namenode-metrics`— Hadoop-Metriken NameNode JMX konfigurieren
- `emr-hadoop-yarn-nodemanager-metrics`— Yarn-Metriken NodeManager JMX konfigurieren
- `emr-hadoop-yarn-resourcemanager-metrics`— ResourceManager JMX Yarn-Metriken konfigurieren
- `emr-hbase-master-metrics`— HBase JMX Master-Metriken konfigurieren
- `emr-hbase-region-server-metrics`— HBase Region JMX Server-Metriken konfigurieren
- `emr-hbase-rest-server-metrics`— HBase REST JMX Server-Metriken konfigurieren
- `emr-hbase-thrift-server-metrics`— HBase Thrift JMX Server-Metriken konfigurieren

In den folgenden Tabellen werden die verfügbaren Eigenschaften und Konfigurationen für alle Klassifizierungen beschrieben.

Eigenschaften von `emr-metrics`

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics_destination</code>	Optional	Legt fest, ob Cluster-Metriken auf Amazon CloudWatch oder Amazon Managed Service for Prometheus veröffentlicht werden.	"CLOUDWATCH"	"CLOUDWATCH", "PROMETHEUS"	Bei dieser Eigenschaft wird nicht zwischen Groß- und Kleinschreibung unterschieden. „Cloudwatch“ ist beispielsweise dasselbe wie " ". CLOUDWATCH
<code>prometheus_endpoint</code>	Optional	Wenn auf "PROMETHEUS" gesetzt <code>metrics_destination</code> ist, konfiguriert diese Eigenschaft den CloudWatch Agenten so, dass er Metriken an den bereitgestellten	N/A	Jeder gültige Amazon Managed Service für Prometheus Remote-Schreibzugriff. URL Das URL Remote-Schreibformat ist <div style="border: 1px solid #ccc; border-radius: 10px; padding: 5px; width: fit-content; margin: 5px auto;"> <pre>https:// aps- worksp aces. <region> s.com/</pre> </div>	Dieses Feld ist erforderlich, wenn <code>metrics_destination</code> es auf "PROMETHEUS" gesetzt ist. Die Bereitstellung schlägt fehl, wenn Sie keinen Schlüssel angeben oder wenn

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
		Remote-Write-Endpunkt von Amazon Managed Service for Prometheus sendet.		workspaces/ <i><workspace_id></i> / api/v1/remote_write	der Wert eine leere Zeichenfolge ist.

emr-system-metrics Eigenschaften

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
metrics_collection_interval	Optional	Wie oft in Sekunden Metriken vom CloudWatch Agenten gesammelt und veröffentlicht werden.	„60“	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Sie können diese Eigenschaft mit der metrics_collection_interval Eigenschaft einzelner Metrikgruppen überschreiben.

emr-system-metrics Konfigurationen

cpu

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der CPU Metriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger CPU Metriken mit oder ohne <code>cpu_</code> Präfix, z. B. <code>usage_active</code> und <code>cpu_time_idle</code> . Gültige Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine CPU Metriken veröffentlicht werden.
<code>metrics_collection_interval</code>	Optional	Wie oft in Sekunden sollte der Agent CPU Metriken sammeln und veröffentlichen.	Der Wert des <code>Globalenmetrics_collection_interval</code> .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für CPU Metriken.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der CPU Metriken, für die keine unaggregierten Metriken veröffentlicht werden sollen.	Keine unaggregierten Metriken veröffentlicht CPU.	Eine durch Kommas getrennte Liste von CPU Metriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle CPU Metriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent Metriken pro Kern veröffentlicht.	<code>"*"</code>	<code>„*“</code> aktiviert Metriken pro Kern. <code>„“</code> deaktiviert Metriken pro Kern.	Der CloudWatch Agent veröffentlicht nur Metriken pro Kern für CPU Metriken, die nicht hinzugefügt wurden. <code>drop_original_metrics</code>

disk

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Festplatten-Metriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger Festplattenmetriken mit oder ohne <code>disk_</code> Präfix, z. B. <code>disk_total</code> und <code>used_perc</code>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Festplattenmetriken veröffentlicht werden.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				ent Gültige Messwerte finden Sie unter Vom CloudWatch Agenten gesammelte Metriken .	
<code>metrics_collection_interval</code>	Optional	Gibt an, wie oft in Sekunden der Agent Festplattenmetriken sammeln und veröffentlichen soll.	Der Wert des <code>globalmetrics_collection_interval</code> .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für Festplattenmesswerte.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der Festplattenmesswerte, für die keine unaggregierten Messwerte veröffentlicht werden sollen.	Es wurden keine unaggregierten Festplattenmetriken veröffentlicht.	Eine durch Kommas getrennte Liste von Festplattenmetriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle Festplattenmetriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent per-mount-point Metriken veröffentlicht.	<code>"*"</code>	„*“ bedeutet alle Einhängpunkte, „“ bedeutet keine Einhängpunkte oder eine durch Kommas getrennte Liste von Bereitstellungs-punkten. z. B. <code>"/, /emr"</code> .	Der CloudWatch Agent veröffentlicht nur per-mount-point Messwerte für Festplattenmetriken, die nicht übernommen wurden. <code>drop_original_metrics</code>

diskio

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Festplatten-E/A-Metriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger Festplatten-IO-Metrikenamen mit oder ohne <code>diskio_</code> Präfix, z. B. <code>diskio_re</code>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Festplatten-IO-Metriken veröffentlicht werden.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				ads und writes Gültige Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.	
metrics_collection_interval	Optional	Gibt an, wie oft in Sekunden der Agent Festplatten-E/A-Metriken sammeln und veröffentlichen soll.	Der Wert des globalenmetrics_collection_interval	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Festplatten-E/A-Metriken.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der Festplatten-IO-Metriken, für die keine unaggregierten Metriken veröffentlicht werden sollen.	Es wurden keine unaggregierten Festplatten-E/A-Metriken veröffentlicht.	Eine durch Kommas getrennte Liste von Festplatten-E/A-Metriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle Festplatten-IO-Metriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent Metriken pro Gerät veröffentlicht.	<code>"*"</code>	„*“ bedeutet alle Speichergereäte, „“ bedeutet keine Speichergereäte oder eine durch Kommas getrennte Liste von Gerätenamen. z. B. <code>"nvme0n1, nvme1n1"</code> .	Der CloudWatch Agent veröffentlicht nur Metriken pro Gerät für Festplatten-E/A-Metriken, die nicht übernommen wurden. <code>drop_original_metrics</code>

mem

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Speichermetriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>mem_</code> Präfix, z. B. <code>mem_available</code> und.	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Speichermetriken veröffentlicht werden.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				available_percent Gültige Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.	
metrics_collection_interval	Optional	Wie oft (in Sekunden) der Agent Speichermetriken sammeln und veröffentlichen soll.	Der Wert des globalen metrics_collection_interval.	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Speichermetriken.

net

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
metrics	Optional	Die Liste der Netzwerkmetriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger Netzwerkmetriken mit oder	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Netzwerkmetriken

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				ohne net_ Präfix, z. B. net_packets_sent und. packets_received Gültige Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.	veröffentlicht werden.
metrics_collection_interval	Optional	Wie oft in Sekunden sollte der Agent Netzwerkmetriken sammeln und veröffentlichen.	Der Wert des Globalenmetrics_collection_interval .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Netzwerkmetriken.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der Netzwerkmetriken, für die keine unaggregierten Messwerte veröffentlicht werden sollen.	Es wurden keine unaggregierten Netzwerkmetriken veröffentlicht.	Eine durch Kommas getrennte Liste von Netzwerkmetriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle Netzwerkmetriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent Metriken pro Schnittstelle veröffentlicht.	"*"	„*“ bedeutet alle Netzwerkschnittstellen, „“ bedeutet keine Netzwerkschnittstellen oder eine durch Kommas getrennte Liste von Schnittstellennamen. z. B. "eth0,eth1" .	Der CloudWatch Agent veröffentlicht nur Schnittstellen-Metriken für Netzwerkmetriken, die nicht hinzugefügt wurden. <code>drop_original_metrics</code>

netstat

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Netzwerkstatistik-Metriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>netstat_</code>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Netzwerkstatistik-Metriken

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				Präfix, z. B. <code>tcp_listen</code> und <code>netstat_udp_socket</code> . Gültige Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken .	veröffentlicht werden.
<code>metrics_collection_interval</code>	Optional	Wie oft in Sekunden sollte der Agent Netzwerkstatistiken sammeln und veröffentlichen.	Der Wert des <code>metrics_collection_interval</code> Globalen.	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für Netzwerkstatistiken.

processes

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Prozessme	Siehe Standardm	Eine durch Kommas	Die Angabe einer leeren

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
		metriken, die der Agent sammeln soll.	Metriken für CloudWatch Agenten bei Amazon EMR	getrennte Liste gültiger Speichermetrikennamen mit oder ohne <code>processes_</code> Präfix, z. B. <code>processes_running</code> und <code>total</code> . Gültige Metriken finden Sie unter <u>Vom CloudWatch Agenten gesammelte Metriken</u>.	Zeichenfolge bedeutet, dass keine Prozessmetriken veröffentlicht werden.
<code>metrics_collection_interval</code>	Optional	Wie oft in Sekunden sollte der Agent Systemprozessmetriken sammeln und veröffentlichen.	Der Wert des <code>Globalenmetrics_collection_interval</code> .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für Systemprozessmetriken.

swap

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Swap-Metriken, die der Agent sammeln soll.	Siehe Standardmetriken für CloudWatch Agenten bei Amazon EMR	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>swap_</code> Präfix, z. B. <code>swap_free</code> und <code>used_percent</code> . Gültige Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Swap-Metriken veröffentlicht werden.
<code>metrics_collection_interval</code>	Optional	Wie oft in Sekunden sollte der Agent Swap-Metriken sammeln und veröffentlichen.	Der Wert des <code>GlobalMetricsCollectionInterval</code> .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für Swap-Metriken.

emr-hadoop-hdfs-datanodeEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch Agent MBean, von dem dieser Agent Metriken sammeln soll, z. B. <code>Hadoop:service=DataNode,name=DataNodeActivity</code> . Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMX YAML Beispieldatensätzen für EMR Amazon-Version 7.0.	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit dem MBean z. B. <code>BlocksCached,BlockRead</code> verknüpft sind.
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Hadoop-Metriken erfasst werden	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
		sollen. DataNode		nden angibt. Akzeptiert nur ganze Zahlen.	

emr-hadoop-hdfs-namenodeEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch Agent MBean, von dem dieser Agent Metriken sammeln soll, z. B. <code>Hadoop:service=NameNode,name=FSNamesystem</code> . Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMX YAML Beispiel	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit dem MBean z. B. <code>BlockCapacityUsedGB</code> .	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			eien für EMR Amazon-Version 7.0.		
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Hadoop-Metriken erfasst werden sollen. NameNode	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

emr-hadoop-yarn-nodemanagerEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch AgentMBean, von dem dieser Agent Metriken sammeln soll, z. B. <code>Hadoop:service=NodeManager,name=NodeManagerMetrics</code>	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit dem verknüpft sind. MBean z. B. <code>MaxCapacity,AllocatedGB</code> .	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMXAMLEispieldateien für EMR Amazon-Version 7.0.		
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Hadoop-Metriken erfasst werden sollen. YARN NodeManager	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

emr-hadoop-yarn-resourcemanagerEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch AgentMBean, von dem dieser Agent	Eine Zeichenfolge, die die kommagetrennte Liste	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
			<p>Metriken sammeln soll, z. B. Hadoop:service=ResourceManager,name=PartitionQueueMetrics Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMX YAML Beispieldateien für EMR Amazon-Version 7.0.</p>	<p>der Metriken enthält, die mit dem verknüpft sind. MBean z. B. MaxCapacity, MaxCapacityVCores .</p>
otel.metric.export.interval	Optional	Wie oft in Millisekunden Hadoop-Metriken erfasst werden sollen. YARN ResourceManager	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.

emr-hbase-master-metrics Eigenschaften

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
<code><custom_bean_name></code>	Optional	N/A	Die PersonMBean, von der der CloudWatch Agent Metriken sammeln soll, wie Hadoop:service=HBase,name=Master,sub=AssignmentManager z. Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMXAMLBearbeitungsbeispieldateien für EMR Amazon-Version 7.0.	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit dem verknüpft sind. MBean z. B. AssignFailedCount, AssignSubmittedCount .
<code>otel.metric.export</code>	Optional	Wie oft in Millisekunden	„60000“	Eine Zeichenfolge

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code>.interval</code>		Master-Metriken erfasst werden sollen. HBase		Die Anzahl der Millisekunden, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

emr-hbase-region-serverEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch Agent MBean, von dem dieser Agent Metriken sammeln soll, z. B. <code>Hadoop:service=HBase,name=RegionServer,sub=IPC</code> . Die MBean Namen der Proben und die entsprechenden Messwerte	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit dem MBean z. B. <code>numActiveHandler,numActivePriorityHandler</code> verknüpft sind.	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			finden Sie in den JMXAMLBeispieldateien für EMR Amazon-Ver sion 7.0.		
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Regionsserver-Metriken erfasst werden sollen. HBase	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

emr-hbase-rest-serverEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch AgentMBean, von dem dieser Agent Metriken sammeln soll, z. B. <code>Hadoop:service=HBase,name=R</code>	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit dem verknüpft sind. MBean z. B.	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			EST Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMX YAML Beispieldat eien für EMR Amazon-Version 7.0.	successfu lPut, succ essfulSca nCount .	
otel.metric.export.interval	Optional	Wie oft in Millisekunden Rest-Server-Metriken erfasst werden sollen. HBase	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

emr-hbase-thrift-serverEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code><custom_bean_name></code>	Optional	N/A	Der CloudWatch AgentMBean, von dem	Eine Zeichenfolge, die die kommagetr	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
			dieser Agent Metriken sammeln soll, z. B. Hadoop:service=HBase,name=ThriftServer,sub=ThriftOn	ennte Liste der Metriken enthält, die mit dem verknüpft sind. MBean z. B. BatchGetMax, BatchGetMean
			Die MBean Namen der Proben und die entsprechenden Messwerte finden Sie in den JMXAMLBespieldaten für EMR Amazon-Version 7.0.	
otel.metric.export.interval	Optional	Wie oft in Millisekunden Thrift-Servermetriken erfasst werden sollen. HBase	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.

Beispiele für Konfigurationen von Systemmetriken

Das folgende Beispiel zeigt, wie der CloudWatch Agent so konfiguriert wird, dass der Export aller Systemmetriken beendet wird.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

Im folgenden Beispiel wird der CloudWatch Agent so konfiguriert, dass er die Standard-Systemmetriken exportiert. Auf diese Weise können Sie den Agenten schnell wieder auf den Export der Standard-Systemmetriken zurücksetzen, wenn Sie die Systemmetriken bereits mindestens einmal neu konfiguriert haben. Durch diesen Reset werden auch alle Anwendungsmetriken entfernt, die zuvor neu konfiguriert wurden.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": []
  }
]
```

Im folgenden Beispiel wird der Cluster so konfiguriert, dass er die `cpumem`, und die `disk` Metriken exportiert.

```
[
  {
    "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": "/",
        "drop_original_metrics": ""
      }
    }
  ]
}
]

```

Die vorherige Beispielkonfiguration hat die folgenden Eigenschaften:

- Alle 30 Sekunden erfasst der Agent die `cpu_guest` Metrik für alle CPUs. Sie finden die aggregierte Metrik unter dem CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name` Namespace.
- Alle 30 Sekunden sammelt der Agent die `cpu_idle` Metrik für alle CPUs. Sie finden die aggregierte Metrik unter dem CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name` Namespace. Der Agent sammelt auch die Metriken pro

CPU. Sie finden sie im selben Namespace. Der Agent sammelt diese Metrik, weil die `drop_original_metrics` Eigenschaft nichts enthält `cpu_idle`, sodass der Agent die Metrik nicht ignoriert.

- Alle 20 Sekunden erfasst der Agent die `mem_active` Metrik. Sie finden die aggregierte Metrik unter dem CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name` Namespace.
- Alle 20 Sekunden erfasst der Agent die `disk_used_percent` Metriken für die `/` und die `/mnt` Festplatten-Mounts. Sie finden die aggregierten Metriken unter dem CloudWatch Namespace. CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name` Der Agent sammelt auch die Metriken pro Mount. Sie finden sie im selben Namespace. Der Agent sammelt diese Metrik, weil die `drop_original_metrics` Eigenschaft nichts enthält `disk_used_percent`, sodass der Agent die Metrik nicht ignoriert.

Beispiele für Konfigurationen von Anwendungsmetriken

Im folgenden Beispiel wird der CloudWatch Agent so konfiguriert, dass der Export von Metriken für den Hadoop Namenode-Dienst beendet wird.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

Im folgenden Beispiel wird ein Cluster für den Export von Hadoop-Anwendungsmetriken konfiguriert.

```
[
  {
    "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": []
    }
  ]
}
]

```

Das vorherige Beispiel hat die folgenden Eigenschaften:

- Alle 20 Sekunden sammelt der Agent die BlockCapacity und CapacityUsedGB -Metriken von Instanzen, auf denen der Hadoop Namenode-Dienst ausgeführt wird.
- Alle 30 Sekunden sammelt der Agent MemNonHeapUsedM Metriken von Instanzen, auf denen der Hadoop Datanode-Dienst ausgeführt wird.
- Alle 30 Sekunden sammelt der Agent die NodeUpdateNumOps Metriken AllocateNumOps und die Metriken von Instanzen, auf denen Hadoop ausgeführt wird. YARN ResourceManager

Beispiel für Amazon Managed Service für Prometheus

Das folgende Beispiel zeigt, wie der CloudWatch Agent so konfiguriert wird, dass er Metriken nach Amazon Managed Service for Prometheus exportiert.

Wenn Sie derzeit Metriken nach Amazon Managed Service for Prometheus exportieren und die Metriken für den Cluster neu konfigurieren und weiterhin Metriken nach Amazon Managed Service for Prometheus exportieren möchten, müssen Sie die Eigenschaften und angeben.

`metrics_destination prometheus_endpoint`

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "prometheus",
      "prometheus_endpoint": "http://amp-workspace/api/v1/remote_write"
    },
    "Configurations": []
  }
]
```

Verwenden Sie das folgende Beispiel, um den CloudWatch Agenten zum Exportieren von Metriken zu CloudWatch verwenden.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "cloudwatch"
    },
    "Configurations": []
  }
]
```

Note

Der CloudWatch Agent verfügt über einen Prometheus-Exporter, der bestimmte Attribute umbenennt. Für die Standardkennzahlen verwendet Amazon Managed Service for Prometheus Unterstriche anstelle der Punkte, die Amazon verwendet. CloudWatch Wenn Sie Amazon Managed Grafana verwenden, um die Standardmetriken in Amazon Managed Service for Prometheus zu visualisieren, werden die Bezeichnungen als `cluster_id`, `instance_id`, `node_type` und angezeigt. `service_name`

CloudWatch Agent für Amazon EMR 7.0.0 konfigurieren

Sie können den CloudWatch Amazon-Agenten so konfigurieren, dass er zusätzliche Systemmetriken verwendet, die über [die standardmäßige CloudWatch Agentenkonfiguration](#) hinausgehen. Die Konfiguration für 7.0.0 erfordert die Verwendung von Bootstrap-Aktionen, für die wir in den folgenden Abschnitten Beispiele bereitgestellt haben. In einer kommenden Version EMR wird Amazon zusätzliche Konfigurationsoptionen über Amazon bereitstellen EMRAPI.

Themen

- [Konfiguration zusätzlicher Systemmetriken mit Amazon EMR 7.0.0](#)
- [Anwendungsmetriken mit Amazon EMR 7.0.0 konfigurieren](#)
- [Amazon Managed Service für Prometheus als Cloud-Speicher für Metriken mit Amazon 7.0.0 konfigurieren EMR](#)

Konfiguration zusätzlicher Systemmetriken mit Amazon EMR 7.0.0

Gehen Sie wie folgt vor, um den Agenten so zu konfigurieren, dass er einen anderen Satz von Systemmetriken in Amazon EMR 7.0.0 verwendet:

1. Erstellen oder wählen Sie einen Bucket in Ihrem Amazon S3 S3-Konto aus, in dem Sie die Konfigurationsdateien speichern möchten, die die CloudWatch Agenten-Metriken spezifizieren.
2. Erstellen Sie die `emr-amazon-cloudwatch-agent.json`-Konfigurationsdatei mit den angegebenen bevorzugten Metriken. Verwenden Sie dazu eine der Methoden, die unter [CloudWatch Agenten-Konfigurationsdatei erstellen](#) beschrieben werden. Weitere Informationen zur Struktur der CloudWatch Agenten-Konfigurationsdatei finden Sie unter [Manuelles Erstellen oder Bearbeiten der CloudWatch Agenten-Konfigurationsdatei](#) im CloudWatch Amazon-Benutzerhandbuch.
3. Navigieren Sie als Nächstes zum `aws-emr-utilities` Repo auf GitHub und laden Sie die folgenden Skripte für Systemmetriken herunter:
 - [install_system_metrics_launcher.sh](#) – Ein Skript, das heruntergeladen und dann `install_system_metrics.sh` im Hintergrund ausführt, sodass der Knoten das Bootstrapping beenden kann.
 - [install_system_metrics.sh](#)— Ein Skript, das darauf wartet, dass die Instanz, auf der es ausgeführt wird, das Bootstrapping abgeschlossen hat, dann die Konfiguration in der Datei herunterlädt und anwendet. JSON

4. Öffnen Sie jede SH-Datei und ersetzen Sie *my-s3-bucket* durch den Namen Ihres Buckets aus Schritt 1.
5. Laden Sie die eine JSON und zwei SH-Dateien in Ihren S3-Bucket hoch.
6. Jetzt können Sie zur EMR Amazon-Konsole navigieren und mit dem CloudWatch Agenten einen neuen Cluster erstellen. Wählen Sie EC2 in der linken Navigationsleiste unter EMR On die Option Cluster und dann Cluster erstellen aus.
7. Wählen Sie im Abschnitt Name und Anwendungen eine EMR Amazon-Version von 7.0.0 oder höher aus.
8. Wählen Sie unter Anwendungspaket das Paket oder die Apps aus, die Sie in Ihrem Cluster installieren möchten, und schließen Sie Amazon CloudWatch Agent in Ihre Auswahl ein.
9. Wählen Sie im Abschnitt Bootstrap actions (Bootstrap-Aktionen) die Option Add (Hinzufügen) aus.
 - Geben Sie für Name **install_system_metrics_launcher.sh** ein.
 - Fügen Sie für Script location (Skript-Standort) **s3://my-s3-bucket/install_system_metrics_launcher.sh** ein. Ersetzen Sie *my-s3-bucket* durch den Pfad zu Ihrem S3-Bucket.
 - Lassen Sie den Argument-Block leer.
10. Wählen Sie Bootstrap-Aktion hinzufügen aus.
11. Fahren Sie mit der Erstellung des Clusters fort, um Ihren Workload-Anforderungen gerecht zu werden.

Wenn Ihr Cluster gestartet wird, veröffentlicht der CloudWatch Agent die Systemmetriken, die Sie in der Konfigurationsdatei angegeben haben, auf CloudWatch.

Anwendungsmetriken mit Amazon EMR 7.0.0 konfigurieren

Sie können den CloudWatch Amazon-Agenten so konfigurieren, dass er Anwendungsmetriken für HDFS und YARN zusätzlich zu Systemmetriken veröffentlicht. Gehen Sie wie folgt vor, um den Agent für die Veröffentlichung von Anwendungsmetriken zu konfigurieren:

1. Erstellen oder wählen Sie einen Bucket in Ihrem Amazon S3 S3-Konto aus, in dem Sie die Konfigurationsdateien speichern möchten, die die CloudWatch Agenten-Metriken spezifizieren.
2. Navigieren Sie als Nächstes zum `aws-emr-utilities` Repo auf GitHub und laden Sie die folgenden Skripte herunter:

- [install_app_metrics_launcher.sh](#) – Ein Skript, das heruntergeladen und dann `install_app_metrics.sh` im Hintergrund ausführt, sodass der Knoten das Bootstrapping beenden kann.
 - [install_app_metrics.sh](#)— Ein Skript, das darauf wartet, dass die Instanz, auf der es ausgeführt wird, das Bootstrapping abgeschlossen hat. Anschließend lädt es die Konfiguration herunter und wendet sie in den YAML Dateien an, die Sie in einem der nächsten Schritte herunterladen werden.
3. Öffnen Sie jede Datei und ersetzen Sie *my-s3-bucket* durch den Namen Ihres Buckets aus Schritt 1.
 4. Laden Sie als Nächstes die folgenden YAML Zuordnungsdateien herunter. Informationen darüber, wie diese YAML Dateien strukturiert sind, finden Sie [javaagent](#) im OpenTelemetry Instrumentation for Java GitHub Repo.
 - [datanode-metrics.yaml](#) – Die Konfiguration für Hadoop DataNode-Metriken.
 - [namenode-metrics.yaml](#) – Die Konfiguration für Hadoop NameNode-Metriken.
 - [nodemanager-metrics.yaml](#) – Die Konfiguration für Yarn NodeManager-Metriken.
 - [resourcemanager-metrics.yaml](#) – Die Konfiguration für Yarn ResourceManager-Metriken.
 5. Laden Sie die zwei SH- und vier YAML Dateien in Ihren S3-Bucket hoch.
 6. Jetzt können Sie zur EMR Amazon-Konsole navigieren und mit dem CloudWatch Agenten einen neuen Cluster erstellen. Wählen Sie EC2 in der linken Navigationsleiste unter EMR On die Option Cluster und dann Cluster erstellen aus.
 7. Wählen Sie im Abschnitt Name und Anwendungen eine EMR Amazon-Version von 7.0.0 oder höher aus.
 8. Wählen Sie unter Anwendungspaket das Paket oder die benutzerdefinierte Gruppe von Apps aus, die Sie in Ihrem Cluster installieren möchten, und schließen Sie CloudWatch Agenten in Ihre Auswahl ein.
 9. Wählen Sie im Abschnitt Bootstrap actions (Bootstrap-Aktionen) die Option Add (Hinzufügen) aus.
 - Geben Sie für Name **install_app_metrics_launcher.sh** ein.
 - Fügen Sie für Script location (Skript-Standort) **s3://my-s3-bucket/install_app_metrics_launcher.sh** ein. Ersetzen Sie *my-s3-bucket* durch den Pfad zu Ihrem S3-Bucket.

- Lassen Sie den Argument-Block leer.
10. Wählen Sie Bootstrap-Aktion hinzufügen aus.
 11. Fahren Sie mit der Erstellung des Clusters fort, um Ihren Workload-Anforderungen gerecht zu werden.

Wenn Ihr Cluster gestartet wird, veröffentlicht der CloudWatch Agent die Anwendungsmetriken, die Sie angegeben haben, zusammen mit den Systemmetriken für CloudWatch.

Amazon Managed Service für Prometheus als Cloud-Speicher für Metriken mit Amazon 7.0.0 konfigurieren EMR

Sie können den CloudWatch Amazon-Agenten so konfigurieren, dass er Metriken stattdessen in Amazon Managed Service for Prometheus veröffentlicht. CloudWatch

Note

Sie können Metriken vom CloudWatch Amazon-Agenten entweder auf Amazon Managed Service for Prometheus oder auf Amazon veröffentlichen CloudWatch, aber Sie können die Metriken nicht für beide Services für denselben Cluster veröffentlichen.

Um den Agenten so zu konfigurieren, dass er Metriken in Amazon Managed Service for Prometheus veröffentlicht, müssen Sie dem EC2 Amazon-Instance-Profil für Amazon die Berechtigung `aps:RemoteWrite` AWS Identity and Access Management (IAM) hinzufügen. EMR Die folgende Beispielrichtlinie enthält die erforderliche Berechtigung:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "aps:RemoteWrite",
      "Resource": "*"
    }
  ]
}
```

Verwenden Sie den CloudWatch Agenten in einem EMR Cluster, um Metriken in Amazon Managed Service for Prometheus zu veröffentlichen

Sobald die Servicerichtlinie über die richtigen Berechtigungen verfügt, führen Sie die folgenden Schritte aus, um einen Cluster zu starten, der den CloudWatch Agenten verwendet, um Metriken in Amazon Managed Service for Prometheus zu veröffentlichen.

1. Verwenden Sie das AWS Management Console oder AWS CLI , um einen Amazon Managed Service for Prometheus Workspace zu erstellen. Weitere Informationen finden Sie unter [Einen Workspace erstellen](#) im Benutzerhandbuch zu Amazon Managed Service für Prometheus.
2. Erstellen Sie einen Bucket in Ihrem Amazon-S3-Konto oder wählen Sie einen aus, in dem Sie die Startdateien speichern möchten, die Amazon Managed Service für Prometheus als Cloud-Speicher angeben.
3. Navigieren Sie als Nächstes zum `aws-emr-utilities` Repo auf GitHub und laden Sie die folgenden Skripts herunter:
 - [add_prometheus_endpoint_launcher.sh](#) – Ein Skript, das heruntergeladen und dann `add_prometheus_endpoint.sh` im Hintergrund ausführt, sodass der Knoten das Bootstrapping beenden kann.
 - [add_prometheus_endpoint.sh](#)— Ein Skript, das darauf wartet, dass die Instance, auf der es ausgeführt wird, den Bootstrapping abgeschlossen hat, und dann den CloudWatch Agenten so konfiguriert, dass er auf dem Amazon Managed Service for Prometheus-Endpoint veröffentlicht, den Sie beim Starten Ihres Clusters als Argument angeben.
4. Öffnen Sie jede Datei und ersetzen Sie `my-s3-bucket` durch den Namen Ihres Buckets aus Schritt 2.
5. Verwenden Sie den, um einen Cluster AWS CLI mit der Bootstrap-Aktion zu erstellenEMR. `add_prometheus_endpoint_launcher.sh` Ersetzen Sie `my-s3-bucket` im folgenden Befehl durch den Bucket, der die Bootstrap-Aktion enthält, und ersetzen Sie `managedpro-remote-write-workspace-url` durch den Remote-Write-Endpoint für Ihren Workspace von Amazon Managed Service für Prometheus. Stellen Sie sicher, dass Sie ein EMR Amazon-Release-Label von `emr-7.0.0` oder höher angeben.

```
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'
```

Wenn Ihr Cluster gestartet wird, veröffentlicht der CloudWatch Agent die gesammelten Metriken in Amazon Managed Service for Prometheus.

Amazon Managed Service für Prometheus als eine Datenquelle für Amazon Managed Grafana verwenden

Sobald Amazon die Cluster-Metriken in Amazon Managed Service for Prometheus veröffentlicht EMR hat, können Sie die Metriken mit Amazon Managed Grafana mit den folgenden Schritten visualisieren:

1. Verwenden Sie den AWS Management Console , um einen Amazon Managed Grafana-Arbeitsbereich und einen Benutzer mit den entsprechenden Berechtigungen zu erstellen. Weitere Informationen finden Sie unter [Einen Workspace erstellen](#) im Benutzerhandbuch zu Amazon Managed Grafana.
2. Fügen Sie Ihren Workspace von Amazon Managed Service für Prometheus als eine Datenquelle zu Amazon Managed Grafana hinzu. Weitere Informationen finden Sie unter [AWS -Datenquellenkonfiguration verwenden, um Amazon Managed Service für Prometheus als Datenquelle hinzuzufügen](#) im Benutzerhandbuch zu Amazon Managed Grafana.

Note

Der CloudWatch Agent verfügt über einen Prometheus-Exporter, der bestimmte Attribute umbenennt. Für die Standardkennzahlen verwendet Amazon Managed Service for Prometheus Unterstriche anstelle der Punkte, die Amazon verwendet. CloudWatch Wenn Sie also Amazon Managed Grafana verwenden, um die Standardmetriken in Amazon Managed Service für Prometheus zu visualisieren, werden die Labels als `jobflow_id`, `instance_id` und `service_name` angezeigt.

Außerdem verwenden alle Anwendungsmetriken, die der CloudWatch Agent auf Amazon Managed Service for Prometheus veröffentlicht, das Label `job` anstelle von `service_name`. Systemmetriken verwenden jedoch weiterhin das Label `service_name`.

Überlegungen und Einschränkungen

- Der native CloudWatch Amazon-Agent ist für Cluster verfügbar, die Sie mit EMR Amazon-Versionen 7.0.0 und höher erstellen.
- Die Konfiguration für den CloudWatch Agenten in Amazon EMR 7.0.0 erfordert die Verwendung von [Bootstrap-Aktionen](#). In einer kommenden Version EMR wird Amazon zusätzliche Konfigurationsoptionen über Amazon bereitstellen EMR API.
- Sie können den EMR CloudWatch Amazon-Agenten nicht installieren, wenn Sie den CloudWatch Agenten bereits mit einer anderen Methode wie AWS Systems Manager Agent (SSMAgent) in der Region bereitgestellt haben, in der Sie Ihren Cluster erstellen. Dies führt dazu, dass Ihr Cluster mit Fehlern beendet wird.

CloudWatch Versionsverlauf des Agenten

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon AmazonCloudWatchAgent enthalten sind EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

AmazonCloudWatchAgent Versionsinformationen

Amazon EMR Release-Etikett	AmazonCloudWatchAgent Ausführung	Komponenten wurden installiert mit AmazonCloudWatchAgent
emr-7.2.0	1,300032,2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager

Amazon EMR Release-Etikett	AmazonCloudWatchAgent Ausführung	Komponenten wurden installiert mit AmazonCloudWatchAgent
		nager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server
emr-7.0.0	1,300031,1	adot-java-agent, emr-amazon-cloudwatch-agent

Delta Lake

Delta Lake ist ein Storage-Layer-Framework für Lakehouse-Architekturen, die üblicherweise auf Amazon S3 basieren. Mit EMR Amazon-Versionen 6.9.0 und höher können Sie [Apache Spark 3.x](#) auf EMR Amazon-Clustern mit Delta Lake-Tabellen verwenden. Weitere Informationen zu Lakehouses mit Delta Lake finden Sie unter <https://delta.io/>.

In der folgenden Tabelle sind die Version von Delta aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Delta EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Delta installiert wurden, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

Delta-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Delta-Version	Mit Delta installierte Komponenten
emr-7.2.0	Delta 3.1.0	Not available.

In der folgenden Tabelle sind die Version von Delta aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Delta EMR installiert.

Die Version der Komponenten, die mit Delta in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Delta-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Delta-Version	Mit Delta installierte Komponenten
emr-6.15.0	Delta 2.4.0	Not available.

Einführung in Delta Lake

Delta Lake ist ein Open-Source-Projekt, das bei der Implementierung moderner Data-Lake-Architekturen hilft, die üblicherweise auf Amazon S3 basieren. Delta Lake bietet folgende Funktionen:

- Atomare, konsistente, isolierte, dauerhafte (ACID) Transaktionen auf Spark. Den Lesern wird während eines Spark-Auftrags eine konsistente Ansicht der Tabelle angezeigt.
- Skalierbares Handling von Metadaten mit verteilter Verarbeitung durch Spark.
- Kombiniert Streaming- und Batch-Anwendungsfälle mit derselben Delta-Tabelle.
- Automatische Durchsetzung von Schemas zur Vermeidung fehlerhafter Datensätze bei der Datenerfassung.
- Zeitreise mit der Versionsverwaltung von Daten.
- Unterstützt Zusammenführungs-, Aktualisierungs- und Löschvorgänge für komplexe Anwendungsfälle wie Change Data Capture (CDC), Streaming-Upserts und mehr.

Einen Cluster verwenden, auf dem Delta Lake installiert ist

Themen

- [Verwenden Sie einen Delta Lake-Cluster mit Flink](#)
- [Einen Delta-Lake-Cluster mit Trino verwenden](#)
- [Verwenden Sie einen Delta-Lake-Cluster mit Spark](#)
- [Einen Delta-Lake-Cluster mit Spark und AWS Glue verwenden](#)

Verwenden Sie einen Delta Lake-Cluster mit Flink

Mit EMR Amazon-Version 6.11 und höher können Sie Delta Lake mit Ihrem Flink-Cluster verwenden. In den folgenden Beispielen wird der verwendet AWS CLI , um mit Delta Lake auf einem Amazon EMR Flink-Cluster zu arbeiten.

Note

Amazon EMR unterstützt den Flink DataStream API, wenn Sie Delta Lake mit einem Flink-Cluster verwenden.

Einen Delta-Lake-Cluster erstellen

1. Erstellen Sie eine Datei, `delta_configurations.json`, mit folgendem Inhalt:

```
[{"Classification": "delta-defaults",  
  "Properties": {"delta.enabled": "true"}}]
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration. Ersetzen Sie `example` Amazon S3 bucket path und subnet ID durch Ihre eigenen Werte.

```
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
```

Eine Flink-Yarn-Sitzung installieren

Um eine Flink-Yarn-Sitzung zu initialisieren, führen Sie den folgenden Befehl aus:

```
flink-yarn-session -d
```

Einen Flink-Auftrag mit Delta Lake erstellen

Die folgenden Beispiele zeigen, wie Sie mit SBT oder Maven für die Erstellung Ihres Flink-Auftrags mit Delta Lake verwenden.

sbt

[sbt](#) ist ein Build-Tool für Scala, das Sie mit wenig bis gar keiner Konfiguration verwenden können, wenn Sie kleine Projekte haben.

```
libraryDependencies += Seq(  
  "io.delta" %% "delta-flink" % deltaConnectorsVersion % "provided",
```

```
"io.delta" %% "delta-standalone" % deltaConnectorsVersion % "provided",
"org.apache.flink" %% "flink-clients" % flinkVersion % "provided",
"org.apache.flink" %% "flink-parquet" % flinkVersion % "provided",
"org.apache.hadoop" % "hadoop-client" % hadoopVersion % "provided",
"org.apache.flink" % "flink-table-common" % flinkVersion % "provided",
"org.apache.flink" %% "flink-table-runtime" % flinkVersion % "provided")
```

Maven

[Maven](#) ist ein Open-Source-Tool zur Build-Automatisierung von der Apache Software Foundation. Mit Maven können Sie einen Flink-Job mit Delta Lake auf Amazon erstellen, veröffentlichen und bereitstellen. 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>

```

Schreiben Sie mit Flink Datastream in eine Delta-Tabelle API

Verwenden Sie das folgende Beispiel, um eine zu erstellen, DeltaSink um in die Tabelle mit einem zu schreiben `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;
}

```

Mit Flink Datastream aus einer Delta-Tabelle lesen API

Verwenden Sie das folgende Beispiel, um eine Grenze zum Lesen aus der Tabelle mit einem DeltaSource zu erstellen `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");
}
```

Sink-Erstellung mit Multi-Cluster-Unterstützung für Delta Lake Standalone

Verwenden Sie das folgende Beispiel, um eine Tabelle DeltaSink zum Schreiben in eine Tabelle mit `A deltaTablePath` - und [Multi-Cluster-Unterstützung](#) zu erstellen:

```
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;
}
```

Den Flink-Auftrag ausführen

Den folgenden Befehl verwenden, um Ihren Job auszuführen:

```
flink run FlinkJob.jar
```

Einen Delta-Lake-Cluster mit Trino verwenden

Mit EMR Amazon-Versionen 6.9.0 und höher können Sie Delta Lake mit Ihrem Trino-Cluster verwenden.

In diesem Tutorial werden wir den Cluster verwenden, AWS CLI um mit Delta Lake auf Amazon EMR Trino zu arbeiten.

Einen Delta-Lake-Cluster erstellen

1. Erstellen Sie eine Datei, `delta_configurations.json`, und legen Sie Werte für den ausgewählten Katalog fest. Wenn Sie beispielsweise den Hive-Metastore als Ihren Katalog verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

Wenn Sie den AWS Glue-Katalog als Ihren Shop verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore":"glue"}}]
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie **example Amazon S3 bucket path** und **subnet ID** durch Ihre eigenen.

```
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
```

Initialisieren Sie die Trino-Sitzung für Delta Lake

Um die Trino-Sitzung zu initialisieren, führen Sie den folgenden Befehl aus

```
trino-cli --catalog delta
```

In eine Delta-Lake-Tabelle schreiben

Erstellen Sie Ihre Tabelle mit den folgenden SQL Befehlen und schreiben Sie in sie:

```
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');
```

Von einer Delta-Lake-Tabelle lesen

Lesen Sie mit dem folgenden SQL Befehl aus Ihrer Tabelle:

```
SELECT * from default.delta_table;
```

Verwenden Sie einen Delta-Lake-Cluster mit Spark

Ab EMR Amazon-Version 6.9.0 können Sie Delta Lake mit Ihrem Spark-Cluster verwenden, ohne dass Bootstrap-Aktionen erforderlich sind. Für EMR Amazon-Versionen 6.8.0 und niedriger können Sie Bootstrap-Aktionen verwenden, um die erforderlichen Abhängigkeiten vorab zu installieren.

In den folgenden Beispielen wird der verwendet AWS CLI , um mit Delta Lake auf einem Amazon EMR Spark-Cluster zu arbeiten.

Um Delta Lake auf Amazon EMR mit dem zu verwenden AWS Command Line Interface, erstellen Sie zunächst einen Cluster. Informationen zur Angabe der Delta Lake-Klassifizierung mit finden Sie unter [Bereitstellen einer Konfiguration mithilfe von AWS Command Line Interface, AWS Command Line Interface wenn Sie einen Cluster erstellen](#) oder [Bereitstellen einer Konfiguration mit Java, SDK wenn Sie einen Cluster erstellen](#).

1. Erstellen Sie eine Datei, `configurations.json`, mit folgendem Inhalt:

```
[{"Classification":"delta-defaults", "Properties":{"delta.enabled":"true"} ]]
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie die Beispiele **bucket path** und **subnet ID** in Amazon S3 durch Ihre eigenen.

```
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
```

Alternativ können Sie einen EMR Amazon-Cluster und eine Spark-Anwendung mit den folgenden Dateien als JAR Abhängigkeiten in einem Spark-Job erstellen:

```
/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

Wenn Sie EMR Amazon-Versionen 6.9.0 oder höher verwenden, verwenden Sie `/usr/share/aws/delta/lib/delta-spark.jar` anstelle von `/usr/share/aws/delta/lib/delta-core.jar`.

Weitere Informationen finden Sie unter [Anwendungen einreichen](#).

Um eine JAR-Abhängigkeit in den Spark-Auftrag aufzunehmen, können Sie der Spark-Anwendung die folgenden Konfigurationseigenschaften hinzufügen:

```
--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"
```

Weitere Informationen zu den Abhängigkeiten von Spark-Aufträgen erhalten Sie unter [Abhängigkeitsmanagement](#).

Wenn Sie EMR Amazon-Versionen 6.9.0 oder höher verwenden, fügen Sie stattdessen die `/usr/share/aws/delta/lib/delta-spark.jar` Konfiguration hinzu.

```
--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"
```

Eine Spark-Sitzung für Delta Lake initialisieren

Die folgenden Beispiele zeigen, wie Sie die interaktive Spark-Shell starten, Spark Submit verwenden oder Amazon EMR Notebooks verwenden, um mit Delta Lake auf Amazon zu arbeiten.

spark-shell

1. Stellen Sie mithilfe von Connect zum Primärknoten her SSH. Weitere Informationen finden Sie unter [Connect mit dem primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.

2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. Um die PySpark Shell zu verwenden, `spark-shell` ersetzen Sie durch `pyspark`.

```
spark-shell \
  --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
  --conf
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

Wenn Sie EMR Amazon-Versionen 6.15.0 oder höher ausführen, müssen Sie außerdem die folgenden Konfigurationen verwenden, um eine differenzierte Zugriffskontrolle auf Basis von Lake Formation mit Delta Lake zu verwenden.

```
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. Stellen Sie mithilfe von Connect zum Primärknoten her SSH. Weitere Informationen finden Sie unter [Connect mit dem primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie den folgenden Befehl ein, um die Spark-Sitzung für Delta Lake zu starten.

```
spark-submit
--conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension"
--conf
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

Wenn Sie EMR Amazon-Versionen 6.15.0 oder höher ausführen, müssen Sie außerdem die folgenden Konfigurationen verwenden, um eine differenzierte Zugriffskontrolle auf Basis von Lake Formation mit Delta Lake zu verwenden.

```
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

Um eine Spark-Sitzung mit Amazon EMR Studio-Notebooks zu initialisieren, konfigurieren Sie Ihre Spark-Sitzung mithilfe des %%configure magischen Befehls in Ihrem EMR Amazon-Notizbuch, wie im folgenden Beispiel. Weitere Informationen finden Sie unter [Use EMR Notebooks Magic](#) im 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"
  }
}
```

Wenn Sie EMR Amazon-Versionen 6.15.0 oder höher ausführen, müssen Sie außerdem die folgenden Konfigurationen verwenden, um eine differenzierte Zugriffskontrolle auf Basis von Lake Formation mit Delta Lake zu verwenden.

```
%%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"
  }
}
```

In eine Delta-Lake-Tabelle schreiben

Das folgende Beispiel zeigt, wie Sie einen Delta Lake-Datensatz erstellen DataFrame und ihn als Delta Lake-Datensatz schreiben. Das Beispiel zeigt, wie Sie mit Datensätzen mit der Spark-Shell arbeiten, während Sie SSH als Standard-Hadoop-Benutzer mit dem primären Knoten verbunden sind.

Note

Um Codebeispiele in die Spark-Shell einzufügen, geben Sie an der Eingabeaufforderung `:paste` ein, fügen das Beispiel ein und drücken dann CTRL + D.

PySpark

Spark umfasst auch eine auf Python basierende Shell, `pyspark`, mit der Sie Prototypen von in Python geschriebenen Spark-Programmen entwickeln können. Rufen Sie genau wie bei `spark-shell`, `pyspark` auf dem Primärknoten auf.

```
## 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");
```

Von einer Delta-Lake-Tabelle lesen

PySpark

```
ddf = spark.table("delta_table")
ddf.show()
```

Scala

```
val ddf = spark.table("delta_table")
ddf.show()
```

SQL

```
SELECT * FROM delta_table;
```

Einen Delta-Lake-Cluster mit Spark und AWS Glue verwenden

Um den AWS Glue-Katalog als Metastore für Delta Lake-Tabellen zu verwenden, erstellen Sie einen Cluster mit den folgenden Schritten. Informationen zur Angabe der Delta Lake-Klassifizierung mithilfe von finden Sie unter [Bereitstellen einer Konfiguration mithilfe von AWS Command Line Interface, AWS Command Line Interface wenn Sie einen Cluster erstellen](#) oder [Bereitstellen einer Konfiguration mithilfe von Java, SDK wenn Sie einen Cluster erstellen](#).

Einen Delta-Lake-Cluster erstellen

1. Erstellen Sie eine Datei, `configurations.json`, mit folgendem Inhalt:

```
[{"Classification":"delta-defaults",
"Properties":{"delta.enabled":"true"}},
{"Classification":"spark-hive-site",
"Properties":
{"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie **example Amazon S3 bucket path** und **subnet ID** durch Ihre eigenen.

```
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
```

Überlegungen und Einschränkungen

- Delta Lake wird für die Verwendung mit EMR Amazon-Versionen 6.9.0 und höher unterstützt. Sie können [Apache Spark 3.x](#) auf EMR Amazon-Clustern mit Delta-Tabellen verwenden.
- Wir empfehlen, das URI s3-Schema für S3-Standortpfade anstelle von s3a zu verwenden, um optimale Leistung, Sicherheit und Zuverlässigkeit zu erzielen. Weitere Informationen finden Sie unter [Arbeiten mit Speicher- und Dateisystemen](#).
- Mit Amazon EMR 7.0 Delta Universal Format (UniForm) und Convert-to-Iceberg werden Kontoauszüge nicht unterstützt.
- Wenn Sie mit Amazon EMR 6.9 und 6.10 Delta Lake-Tabellendaten in Amazon S3 speichern, werden die Spaltendaten NULL nach dem Umbenennen der Spalten gespeichert. Dieses Problem wurde mit Amazon EMR 6.11 behoben. Weitere Informationen zum experimentellen Vorgang zum Umbenennen von Spalten finden Sie unter [Vorgang zum Umbenennen von Spalten](#) im Delta-Lake-Benutzerhandbuch.
- Wenn Sie außerhalb von Apache Spark eine Datenbank im AWS Glue-Datenkatalog erstellen, könnte die Datenbank ein leeres LOCATION Feld enthalten. Da Spark das Erstellen von Datenbanken mit einer leeren Standorteigenschaft nicht zulässt, erhalten Sie die folgende Fehlermeldung, wenn Sie Spark in Amazon verwenden, EMR um eine Delta-Tabelle in einer Glue-Datenbank zu erstellen und die Datenbank eine leere LOCATION Eigenschaft hat:

```
IllegalArgumentException: Can not create a Path from an empty string
```

Um dieses Problem zu beheben, erstellen Sie die Datenbank im Datenkatalog mit einem gültigen, nicht leeren Pfad für das LOCATION-Feld. Schritte zur Implementierung dieser Lösung finden Sie unter [Unzulässige Argumentausnahme beim Erstellen einer Tabelle](#) im Amazon-Athena-Benutzerhandbuch.

Versionsverlauf von Delta

In der folgenden Tabelle sind die Delta-Versionen aufgeführt, die in jeder Release-Version von Amazon enthalten sind EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Delta Versionsinformationen

Amazon EMR Release-Etikett	Delta-Version	Mit Delta installierte Komponenten
emr-7.2.0	3.1.0	Not available.
emr-7.1.0	3.0.0	Not available.
emr-7.0.0	3.0.0	Not available.
emr-6.15.0	2.4.0	Not available.
emr-6.14.0	2.4.0	Not available.
emr-6.13.0	2.4.0	Not available.
emr-6.12.0	2.4.0	Not available.
emr-6.11.1	2.2.0	Not available.
emr-6.11.0	2.2.0	Not available.
emr-6.10.1	2.2.0	Not available.
emr-6.10.0	2.2.0	Not available.
emr-6.9.1	2.1.0	Not available.
emr-6.9.0	2.1.0	Not available.

Apache Flink

[Apache Flink](#) ist eine Streaming-Datenfluss-Engine, die mittels Datenquellen mit hohem Durchsatz die Echtzeitverarbeitung von Streams ermöglicht. Flink unterstützt Ereigniszeitsemantik für out-of-order Ereignisse, Exactly-Once-Semantik und Gegendruckkontrolle und ist für das Schreiben von Streaming- und Batch-Anwendungen optimiert. APIs

Außerdem verfügt Flink über Konnektoren für Datenquellen von Drittanbieter, wie z. B. die folgenden:

- [Amazon Kinesis Data Streams](#)
- [Apache Kafka](#)
- [Flink-Elasticsearch-Konnektor](#)
- [Twitter-Streaming API](#)
- [Cassandra](#)

Amazon EMR unterstützt Flink als YARN Anwendung, sodass Sie Ressourcen zusammen mit anderen Anwendungen innerhalb eines Clusters verwalten können. Flink-on — YARN ermöglicht es Ihnen, vorübergehende Flink-Jobs einzureichen, oder Sie können einen Cluster mit langer Laufzeit erstellen, der mehrere Jobs akzeptiert und Ressourcen entsprechend der Gesamtreservierung zuweist. YARN

Flink ist in den EMR Amazon-Release-Versionen 5.1.0 und höher enthalten.

Note

Support für die `FlinkKinesisConsumer` Klasse wurde in der EMR Amazon-Release-Version 5.2.1 hinzugefügt.

In der folgenden Tabelle sind die Version von Flink aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Flink EMR installiert.

[Die Version der Komponenten, die in dieser Version mit Flink installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Flink-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Flink-Version	Mit Flink installierte Komponenten
emr-7.2.0	Flink 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, flink-jobmanager-config, hudi, delta

In der folgenden Tabelle sind die Version von Flink aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Flink EMR installiert.

Die Version der Komponenten, die mit Flink in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Flink-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Flink-Version	Mit Flink installierte Komponenten
emr-6.15.0	Flink 1.17.1	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,

EMRAmazon-Freigabeetikett	Flink-Version	Mit Flink installierte Komponenten
		flink-jobmanager-config, hudi, delta-standalone-connectors

In der folgenden Tabelle sind die Version von Flink aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Flink EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Flink installiert wurden, finden Sie unter Komponentenversionen der Version 5.36.2.](#)

Flink-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Flink-Version	Mit Flink installierte Komponenten
emr-5.36.2	Flink 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

Themen

- [Erstellen eines Clusters mit Flink](#)
- [Konfiguration von Flink in Amazon EMR](#)
- [Arbeiten mit Flink-Jobs in Amazon EMR](#)
- [Verwenden der Scala-Shell](#)
- [Suchen der Flink Webschnittstelle](#)
- [Flink Autoscaler](#)

- [Optimieren der Neustartzeiten von Aufträgen für die Aufgabenwiederherstellung und -skalierung](#)
- [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#)
- [Versionsverlauf von Flink](#)

Erstellen eines Clusters mit Flink

Sie können einen Cluster mit dem AWS Management Console AWS CLI, oder einem starten AWS SDK.

So starten Sie einen Cluster mit Flink über die Konsole

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie für Softwarekonfiguration EMR Release emr-5.1.0 oder höher aus.
4. Wählen Sie Flink als Anwendung zusammen mit anderen zu installierenden Anwendungen aus.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Um einen Cluster mit Flink von der AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl:

```
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

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Konfiguration von Flink in Amazon EMR

Konfigurieren Sie Flink mit Hive Metastore und Glue Catalog

EMRAmazon-Versionen 6.9.0 und höher unterstützen sowohl Hive Metastore als auch AWS Glue Catalog mit dem Apache Flink-Konnektor zu Hive. In diesem Abschnitt werden die Schritte beschrieben, die zur Konfiguration von [AWS Glue Catalog](#) und [Hive Metastore](#) mit Flink erforderlich sind.

Themen

- [Verwenden Sie den Hive Metastore](#)
- [Verwenden Sie den AWS Glue-Datenkatalog](#)

Verwenden Sie den Hive Metastore

1. Erstellen Sie einen EMR Cluster mit Version 6.9.0 oder höher und mindestens zwei Anwendungen: Hive und Flink.
2. Verwenden Sie [Script Runner](#), um das folgende Skript als Schrittfunktion auszuführen:

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

Verwenden Sie den AWS Glue-Datenkatalog

1. Erstellen Sie einen EMR Cluster mit Version 6.9.0 oder höher und mindestens zwei Anwendungen: Hive und Flink.
2. Wählen Sie in den Einstellungen des AWS Glue Data Catalogs die Option Für Hive-Tabellenmetadaten verwenden aus, um den Data Catalog im Cluster zu aktivieren.
3. Verwenden Sie [Script Runner](#), um das folgende Skript als Schrittfunktion [auszuführen: Befehle und Skripts auf einem EMR Amazon-Cluster](#) ausführen:

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

Konfigurieren Sie Flink mit einer Konfigurationsdatei

Sie können die EMR Amazon-Konfiguration verwenden API, um Flink mit einer Konfigurationsdatei zu konfigurieren. Die Dateien, die innerhalb von konfiguriert werden können, API sind:

- `flink-conf.yaml`
- `log4j.properties`
- `flink-log4j-session`
- `log4j-cli.properties`

Die Hauptkonfigurationsdatei für Flink heißt `flink-conf.yaml`.

So konfigurieren Sie die Anzahl der Aufgaben-Slots für Flink mithilfe der AWS CLI

1. Erstellen Sie eine Datei, `configurations.json`, mit folgendem Inhalt:

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "taskmanager.numberOfTaskSlots": "2"
    }
  }
]
```

]

2. Im nächsten Schritt erstellen Sie einen Cluster mit der folgenden Konfiguration:

```
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

Sie können einige Konfigurationen auch mit dem Flink API ändern. Weitere Informationen finden Sie unter [Konzepte](#) in der Flink-Dokumentation.

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Dazu verwenden Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Parallelitätsoptionen

Als Eigentümer Ihrer Anwendung wissen Sie am besten, welche Ressourcen Aufgaben innerhalb von Flink zugewiesen werden müssen. Für Beispiele in dieser Dokumentation verwenden Sie die gleiche Anzahl von Aufgaben wie die Aufgaben-Instances, die Sie für die Anwendung nutzen. Wir empfehlen diese Vorgehensweise generell für die anfängliche Parallelität. Sie können jedoch die Granularität der Parallelität mithilfe von Aufgaben-Slots erhöhen. Dabei sollte die Anzahl von [virtuellen Cores](#) pro Instance im Allgemeinen nicht überschritten werden. Weitere Informationen über die Architektur von Flink finden Sie unter [Konzepte](#) in der Flink-Dokumentation.

Konfiguration von Flink auf einem EMR Cluster mit mehreren Primärknoten

Der JobManager von Flink bleibt während des Failover-Prozesses für den primären Knoten in einem EMR Amazon-Cluster mit mehreren Primärknoten verfügbar. Ab Amazon EMR 5.28.0 ist JobManager Hochverfügbarkeit auch automatisch aktiviert. Es ist keine manuelle Konfiguration erforderlich.

Bei EMR Amazon-Versionen 5.27.0 oder früher JobManager handelt es sich um einen einzigen Fehlerpunkt. Wenn der JobManager fehlschlägt, gehen alle Jobstatus verloren und die laufenden Jobs werden nicht wieder aufgenommen. Sie können JobManager Hochverfügbarkeit aktivieren, indem Sie die Anzahl der Anwendungsversuche, das Checkpointing und die Aktivierung ZooKeeper als Statusspeicher für Flink konfigurieren, wie das folgende Beispiel zeigt:

```
[
  {
    "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"
    }
  }
]
```

Sie müssen sowohl die maximale Anzahl der Anwendungsmasterversuche für Flink als auch die maximalen Anwendungsversuche für YARN Flink konfigurieren. Weitere Informationen finden Sie unter [Konfiguration der Hochverfügbarkeit von YARN Clustern](#). Möglicherweise möchten Sie auch das Flink-Checkpointing so konfigurieren, dass neu gestartete Jobs von zuvor abgeschlossenen Checkpoints JobManager wiederhergestellt werden. Weitere Informationen finden Sie unter [Flink-Checkpointing](#).

Konfiguration der Größe des Speicherprozesses

Für EMR Amazon-Versionen, die Flink 1.11.x verwenden, müssen Sie die Gesamtspeicherprozessgröße sowohl für () als auch für JobManager (`jobmanager.memory.process.size`) in konfigurieren. TaskManager `taskmanager.memory.process.size` `flink-conf.yaml` Sie können diese Werte festlegen, indem Sie entweder den Cluster mit der Konfiguration konfigurieren API oder diese Felder manuell über auskommentieren. SSH Flink bietet die folgenden Standardwerte.

- `jobmanager.memory.process.size`: 1 600 m
- `taskmanager.memory.process.size`: 1 728 m

Um JVM Metaspace und Overhead auszuschließen, verwenden Sie stattdessen die gesamte Flink-Speichergröße (). `taskmanager.memory.flink.size` `taskmanager.memory.process.size` Der Standardwert von `taskmanager.memory.process.size` beträgt 1 280 m. Es wird nicht empfohlen, sowohl `taskmanager.memory.process.size` als auch `taskmanager.memory.process.size` zu setzen.

Alle EMR Amazon-Versionen, die Flink 1.12.0 und höher verwenden, haben die im Open-Source-Set für Flink aufgeführten Standardwerte als Standardwerte bei AmazonEMR, sodass Sie sie nicht selbst konfigurieren müssen.

Größe der Protokollausgabedatei konfigurieren

Flink-Anwendungscontainer erstellen drei Arten von Protokolldateien und schreiben in sie: `.out`-Dateien, `.log`-Dateien und `.err`-Dateien. Nur `.err`-Dateien werden komprimiert und aus dem Dateisystem entfernt, während `.log`- und `.out`-Protokolldateien im Dateisystem verbleiben. Um sicherzustellen, dass diese Ausgabedateien verwaltbar bleiben und der Cluster stabil bleibt, können Sie die Protokollrotation in `log4j.properties` so konfigurieren, dass eine maximale Anzahl von Dateien festgelegt und deren Größe begrenzt wird.

EMRAmazon-Versionen 5.30.0 und höher

Ab Amazon EMR 5.30.0 verwendet Flink das Logging-Framework `log4j2` mit dem Namen der Konfigurationsklassifikation. Die folgende Beispielkonfiguration demonstriert `flink-log4j`. das `log4j2`-Format.

```
[
```

```

{
  "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"
  },
}
]

```

EMRAmazon-Versionen 5.29.0 und früher

In den EMR Amazon-Versionen 5.29.0 und früher verwendet Flink das Logging-Framework log4j. Die folgende Beispielkonfiguration veranschaulicht das log4j Format.

```

[
  {
    "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"
    },
  }
]

```

Flink für die Ausführung mit Java 11 konfigurieren

EMRAmazon-Versionen 6.12.0 und höher bieten Java 11-Laufzeitunterstützung für Flink. In den folgenden Abschnitten wird beschrieben, wie der Cluster so konfiguriert wird, dass er Java-11-Laufzeitunterstützung für Flink bereitstellt.

Themen

- [Konfigurieren Sie Flink für Java 11, wenn Sie einen Cluster erstellen](#)
- [Flink für Java 11 auf einem laufenden Cluster konfigurieren](#)
- [Bestätigen Sie die Java-Laufzeit für Flink auf einem laufenden Cluster](#)

Konfigurieren Sie Flink für Java 11, wenn Sie einen Cluster erstellen

Gehen Sie wie folgt vor, um einen EMR Cluster mit Flink und Java 11-Runtime zu erstellen. Die Konfigurationsdatei, in der Sie die Java-11-Laufzeitunterstützung hinzufügen, befindet sich. `flink-conf.yaml`

Console

Um einen Cluster mit Flink und Java 11-Laufzeit in der Konsole zu erstellen

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie EC2 im Navigationsbereich EMRunter „Cluster“ und anschließend „Cluster erstellen“ aus.
3. Wählen Sie EMR Amazon-Version 6.12.0 oder höher und installieren Sie die Flink-Anwendung. Wählen Sie alle anderen Anwendungen aus, die Sie auf Ihrem Cluster installieren möchten.
4. Fahren Sie mit der Einrichtung Ihres Clusters fort. Verwenden Sie im Bereich optionale Softwareeinstellungen die Standardoption Konfiguration eingeben und geben Sie die folgende Konfiguration ein:

```
[
  {
    "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. Fahren Sie mit der Einrichtung und dem Start Ihres Clusters fort.

AWS CLI

Um einen Cluster mit Flink und Java 11-Runtime aus dem zu erstellen CLI

1. Erstellen Sie eine Konfigurationsdatei `configurations.json`, die Flink für die Verwendung von Java 11 konfiguriert.

```

[
  {
    "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. Erstellen Sie aus dem AWS CLI einen neuen EMR Cluster mit EMR Amazon-Version 6.12.0 oder höher und installieren Sie die Flink-Anwendung, wie im folgenden Beispiel gezeigt:

```

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

```

Flink für Java 11 auf einem laufenden Cluster konfigurieren

Gehen Sie wie folgt vor, um einen laufenden EMR Cluster mit Flink und Java 11-Runtime zu aktualisieren. Die Konfigurationsdatei, in der Sie die Java-11-Laufzeitunterstützung hinzufügen, befindet sich. `flink-conf.yaml`

Console

Um einen laufenden Cluster mit Flink und Java 11-Runtime in der Konsole zu aktualisieren

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie EC2 im Navigationsbereich EMR unter „Ein“ die Option „Cluster“ und wählen Sie dann den Cluster aus, den Sie aktualisieren möchten.

Note

Der Cluster muss Amazon EMR Release 6.12.0 oder höher verwenden, um Java 11 zu unterstützen.

3. Wählen Sie die Registerkarte Konfigurationen aus.
4. Wählen Sie im Abschnitt Instance-Gruppenkonfigurationen die Instance-Gruppe Running aus, die Sie aktualisieren möchten, und wählen Sie dann Neukonfiguration aus dem Menü mit den Listenaktionen aus.
5. Konfigurieren Sie die Instance-Gruppe mit der Option Attribute bearbeiten wie folgt neu. Wählen Sie nach jeder Konfiguration Neue Konfiguration hinzufügen aus.

Klassifizierung	Eigenschaft	Wert
flink-conf	containerized.taskmanager.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	containerized.master.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	env.java.home	/usr/lib/jvm/jre-11

6. Wählen Sie Änderungen speichern aus um die Konfigurationen hinzuzufügen. .

AWS CLI

Um einen laufenden Cluster so zu aktualisieren, dass er Flink und Java 11-Runtime von der CLI

Verwenden Sie den Befehl `modify-instance-groups`, um eine neue Konfiguration für eine Instance-Gruppe in einem laufenden Cluster anzugeben.

1. Erstellen Sie eine Konfigurationsdatei `configurations.json`, die Flink für die Verwendung von Java 11 konfiguriert. Ersetzen Sie im folgenden Beispiel `ig-1xxxxxxx9` durch die ID für die Instanzgruppe, die Sie neu konfigurieren möchten. Speichern Sie die Datei im selben Verzeichnis, in dem Sie den `modify-instance-groups`-Befehl ausführen werden.

```
[
  {
    "InstanceGroupId": "ig-1xxxxxxx9",
    "Configurations": [
      {
        "Classification": "flink-conf",
        "Properties": {
          "containerized.taskmanager.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "containerized.master.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "env.java.home": "/usr/lib/jvm/jre-11"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. Führen Sie von der AWS CLI aus den folgenden Befehl aus. Ersetzen Sie die ID für die Instance-Gruppe, die Sie neu konfigurieren möchten:

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```

Bestätigen Sie die Java-Laufzeit für Flink auf einem laufenden Cluster

Um die Java-Laufzeit für einen laufenden Cluster zu ermitteln, melden Sie sich beim primären Knoten an, SSH wie unter [Connect zum Primärknoten herstellen mit](#) beschrieben SSH. Führen Sie anschließend den folgenden Befehl aus:

```
ps -ef | grep flink
```

Der `ps`-Befehl mit der `-ef`-Option listet alle laufenden Prozesse auf dem System auf. Sie können diese Ausgabe mit `grep` filtern, um Erwähnungen der Zeichenfolge `flink` zu finden. Überprüfen Sie die Ausgabe für den Wert Java Runtime Environment (JRE), `jre-XX`. In der folgenden Ausgabe wird `jre-11` angegeben, dass Java 11 zur Laufzeit von Flink abgerufen wird.

```
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.
```

Alternativ können Sie [sich mit dem Befehl `flink-yarn-session -d` beim primären Knoten anmelden SSH](#) und eine YARN Flink-Sitzung starten. Die Ausgabe zeigt die Java Virtual Machine (JVM) für Flink, `java-11-amazon-corretto` im folgenden Beispiel:

```
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
```

Arbeiten mit Flink-Jobs in Amazon EMR

Es gibt mehrere Möglichkeiten, mit Flink in Amazon EMR zu interagieren: über die Konsole, die Flink-Schnittstelle in der Resource Manager Tracking-Benutzeroberfläche und in der Befehlszeile. Mit jedem dieser Programme können Sie eine JAR-Datei an eine Flink-Anwendung senden. Sobald Sie eine JAR-Datei eingereicht haben, wird sie zu einem Auftrag, der von Flink verwaltet wird JobManager. Die JobManager befindet sich auf dem YARN-Knoten, der den Application Master-Daemon der Flink-Sitzung hostet.

Sie können eine Flink-Anwendung als YARN-Auftrag auf einem Cluster mit langer Laufzeit oder auf einem vorübergehenden Cluster ausführen. In einem Cluster mit langer Laufzeit können Sie mehrere Flink-Aufträge an einen Flink-Cluster senden, der auf Amazon EMR ausgeführt wird. Wenn Sie einen Flink-Auftrag auf einem vorübergehenden Cluster ausführen, existiert Ihr Amazon-EMR-Cluster nur für die Zeit, die zum Ausführen der Flink-Anwendung benötigt wird, sodass Ihnen nur die verbrauchten Ressourcen und die Zeit in Rechnung gestellt werden. Sie können einen Flink-Auftrag mit der Amazon-EMR-AddSteps-API-Operation als Schrittargument für die RunJobFlow-Operation und über die Befehle AWS CLI, `add-steps` oder `create-cluster` übermitteln.

Eine Flink-YARN-Anwendung als Schritt auf einem Cluster mit langer Laufzeit starten

Um eine Flink-Anwendung zu starten, an die mehrere Clients über YARN-API-Operationen Arbeit einreichen können, müssen Sie entweder einen Cluster erstellen oder eine Flink-Anwendung zu einem vorhandenen Cluster hinzufügen. Eine Anleitung zur Erstellung eines neuen Clusters finden Sie unter [Erstellen eines Clusters mit Flink](#). Um eine YARN-Sitzung auf einem vorhandenen Cluster zu starten, führen Sie die folgenden Schritte über die Konsole, AWS CLI oder das Java-SDK aus.

Note

Der Befehl `flink-yarn-session` wurde in Amazon-EMR-Version 5.5.0 als Wrapper für das Skript `yarn-session.sh` zur Vereinfachung der Ausführung hinzugefügt. Wenn Sie eine frühere Version von Amazon EMR verwenden, ersetzen Sie `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` mit Argumente in der Konsole oder `Args` im AWS CLI-Befehl.

So senden Sie einen Flink-Auftrag auf einem vorhandenen Cluster von der Konsole aus

Senden Sie die Flink-Sitzung mit dem Befehl `flink-yarn-session` in einem vorhandenen Cluster.

1. Öffnen Sie die Amazon-EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie in der Cluster-Liste den Cluster aus, den Sie zuvor gestartet haben.
3. Wählen Sie auf der Cluster-Detailseite **Steps (Schritte)** und **Add Step (Schritt hinzufügen)** aus.
4. Befolgen Sie die folgenden Richtlinien, um die Parameter einzugeben, und wählen Sie dann **Hinzufügen** aus.

Parameter	Beschreibung
Step type (Schritttyp)	Custom JAR
Name	Ein Name zur Identifizierung des Schritts. Zum Beispiel <i><example-flink-step-name></i> .
Jar location (Jar-Ort)	command-runner.jar
Argumente	Der Befehl <code>flink-yarn-session</code> mit für Ihre Anwendung geeigneten Argumenten. Beispielsweise startet <i>flink-yarn-session -d</i> eine Flink-Sitzung in Ihrem YARN-Cluster in einem getrennten Zustand (-d). Weitere Informationen finden Sie unter YARN-Einrichtung in der aktuellen Flink-Dokumentation für Argumentdetails.

So senden Sie einen Flink-Auftrag auf einem vorhandenen Cluster mit der AWS CLI aus

- Verwenden Sie den `add-steps`-Befehl, um einem Cluster mit langer Laufzeit einen Flink-Auftrag hinzuzufügen. Der folgende Beispielbefehl gibt `Args="flink-yarn-session", "-d"` an, sodass eine Flink-Sitzung innerhalb Ihres YARN-Clusters in einem getrennten Zustand (-d) gestartet werden soll. Weitere Informationen finden Sie unter [YARN-Einrichtung](#) in der aktuellen Flink-Dokumentation für Argumentdetails.

```
aws emr add-steps --cluster-id <j-XXXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-flink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

Senden Sie Ihre Arbeit an eine bestehende Flink-Anwendung auf einem Cluster mit langer Laufzeit

Wenn Sie bereits eine bestehende Flink-Anwendung auf einem Cluster mit langer Laufzeit haben, können Sie die Flink-Anwendungs-ID des Clusters angeben, um Arbeit an diesen zu senden. Um die

Anwendungs-ID zu erhalten, führen Sie `yarn application -list` auf der AWS CLI oder über die [YarnClient](#) -API-Operation aus:

```
$ 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
```

Die Anwendungs-ID für diese Flink-Sitzung lautet `application_1473169569237_0002`, mit der Sie Arbeit aus dem AWS CLI oder einem SDK an die Anwendung senden können.

Example SDK für 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>
```

Senden eines kurzlebigen Flink-Auftrags

Die folgenden Beispiele starten einen vorübergehenden Cluster, der einen Flink-Auftrag ausführt und dann nach Abschluss beendet wird.

Example SDK für 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\n"
                "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

Verwenden Sie den Unterbefehl `create-cluster`, um einen kurzlebigen Cluster zu erstellen, der beendet wird, wenn der Flink-Auftrag abgeschlossen ist:

```
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/\""
```

Verwenden der Scala-Shell

Die Flink-Scala-Shell für EMR-Cluster ist so konfiguriert, dass nur neue YARN Sitzungen starten. Sie können die Scala-Shell wie im Folgenden beschrieben verwenden.

Den Flink Scala-Shell auf dem Primärknoten verwenden

1. Melden Sie sich mit SSH beim Primärknoten an, wie unter [SSH mit dem Primärknoten verbinden](#) beschrieben.
2. Geben Sie Folgendes ein, um eine Shell zu starten:

In Amazon-EMR-Version 5.5.0 und höher können Sie den folgenden Befehl verwenden, um einen Yarn-Cluster für die Scala-Shell mit einem TaskManager zu starten.

```
% flink-scala-shell yarn 1
```

In früheren Versionen von Amazon EMR verwenden Sie:

```
% /usr/lib/flink/bin/start-scala-shell.sh yarn 1
```

Damit wird die Flink Scala-Shell gestartet, sodass Sie Flink interaktiv nutzen können. Wie bei anderen Schnittstellen und Optionen können Sie im Beispiel verwendeten Optionswert `-n` basierend auf der Anzahl der Aufgaben, die Sie in der Shell ausführen möchten, skalieren.

Weitere Informationen finden Sie unter [Scala REPL](#) in der offiziellen Apache-Flink-Dokumentation.

Suchen der Flink Webschnittstelle

Der Application Master, der zur Flink-Anwendung gehört, hostet das Flink-Webinterface. Dies ist eine alternative Möglichkeit, ein JAR als Auftrag einzureichen oder den aktuellen Status anderer Aufträge einzusehen. Die Flink-Webschnittstelle ist aktiv, solange eine Flink-Sitzung ausgeführt wird. Wenn bereits ein lang laufender YARN-Auftrag aktiv ist, können Sie die Anweisungen im Thema [Verbindung zum Primärknoten mit SSH](#) im Verwaltungshandbuch für Amazon EMR befolgen, um sich mit dem YARN ResourceManager zu verbinden. Wenn Sie beispielsweise einen SSH-Tunnel eingerichtet und in Ihrem Browser einen Proxy aktiviert haben, wählen Sie die ResourceManager-Verbindung unter Verbindungen auf der Cluster-Detailseite Ihres EMR-Clusters aus.

Cluster: Development Cluster Waiting Cluster ready after last step completed.

Connections:



[Resource Manager ... \(View All\)](#)

Nachdem Sie den ResourceManager gefunden haben, wählen Sie die YARN-Anwendung aus, die eine Flink-Sitzung hostet. Wählen Sie den Link unter der Spalte Tracking UI aus.

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	<div style="width: 100px; height: 10px; background-color: #ccc;"></div>	ApplicationMaster

First Previous

In der Flink-Webschnittstelle können Sie die Konfiguration ansehen, Ihre eigene benutzerdefinierte JAR-Datei als Auftrag senden oder in Bearbeitung befindliche Aufträge überwachen.

Apache Flink Dashboard Overview Version: 1.1.1

Task Managers: 1

Task Slots: 1

Available Task Slots: 1

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
------------	----------	----------	----------	--------	-------	--------

Flink Autoscaler

Übersicht

Amazon-EMR-Versionen 6.15.0 und höher unterstützen Flink Autoscaler. Die Auftrag-Autoscaler-Funktion sammelt Metriken von laufenden Flink-Streaming-Aufträgen und skaliert automatisch die einzelnen Scheitelpunkte der Aufträge. Dadurch wird der Gegendruck reduziert und das von Ihnen festgelegte Nutzungsziel wird erfüllt.

Weitere Informationen finden Sie im Abschnitt [Autoscaler](#) in der Dokumentation zu Apache Flink Kubernetes Operator.

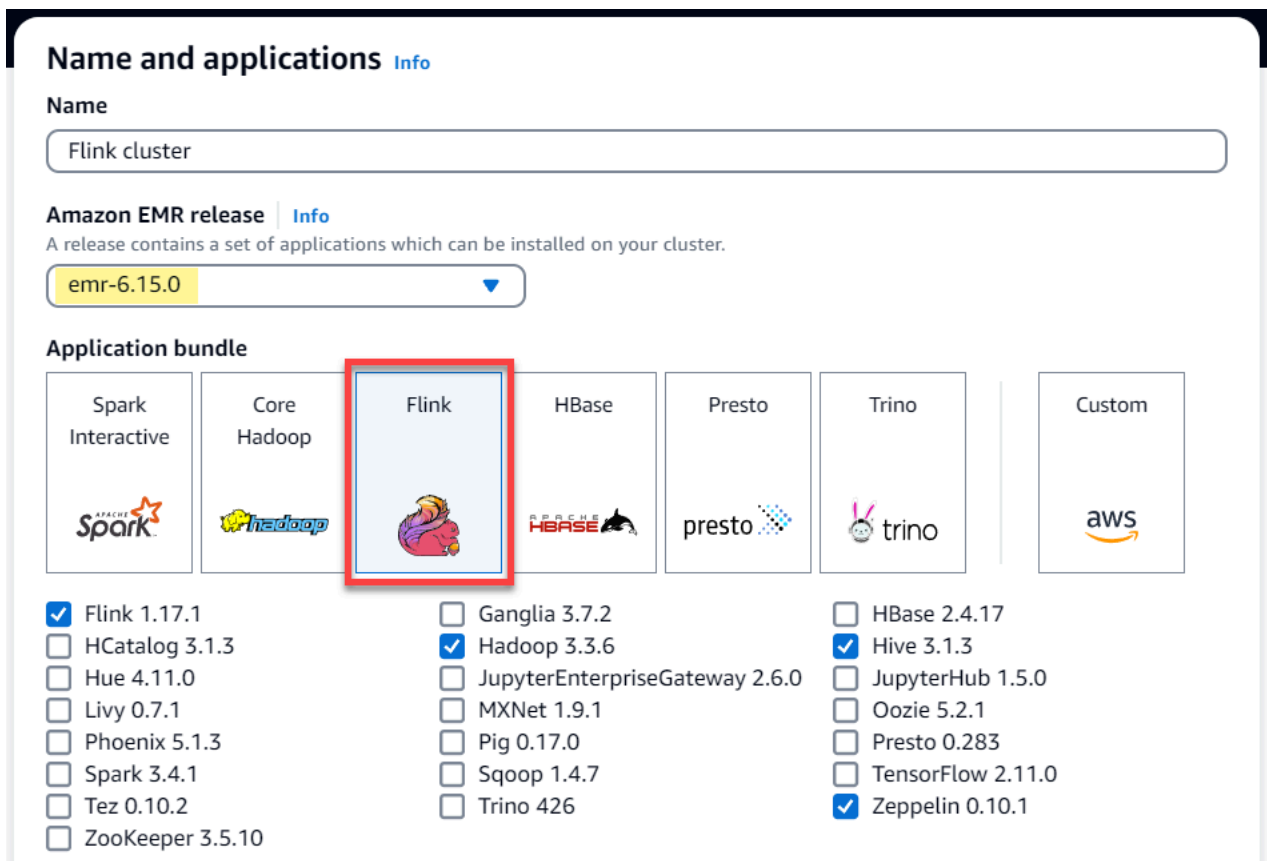
Überlegungen

- Flink Autoscaler wird mit Amazon EMR 6.15.0 und höher unterstützt.
- Flink Autoscaler wird nur für Streaming-Aufträge unterstützt.
- Nur der adaptive Scheduler wird unterstützt. Der Standard-Scheduler wird nicht unterstützt.
- Wir empfehlen Ihnen, die Cluster-Skalierung zu aktivieren, um eine dynamische Ressourcenbereitstellung zu ermöglichen. Amazon EMR Managed Scaling wird bevorzugt, weil die Metrikauswertung alle 5 bis 10 Sekunden erfolgt. In diesem Intervall kann sich Ihr Cluster leichter an die Änderung der erforderlichen Cluster-Ressourcen anpassen.

Autoscaler aktivieren

Gehen Sie wie folgt vor, um den Flink Autoscaler zu aktivieren, wenn Sie einen Amazon EMR in einem EC2-Cluster erstellen.

1. Erstellen Sie in der Amazon-EMR-Konsole einen neuen EMR-Cluster:
 - a. Wählen Sie Amazon EMR Version `emr-6.15.0` oder höher aus. Wählen Sie das Flink-Anwendungspaket und alle anderen Anwendungen aus, die Sie möglicherweise in Ihren Cluster aufnehmen möchten.



Name and applications [Info](#)

Name

Flink cluster

Amazon EMR release [Info](#)

A release contains a set of applications which can be installed on your cluster.

emr-6.15.0

Application bundle

Spark Interactive Core Hadoop **Flink** HBase Presto Trino Custom

Flink 1.17.1 HCatalog 3.1.3 Hue 4.11.0 Livy 0.7.1 Phoenix 5.1.3 Spark 3.4.1 Tez 0.10.2 ZooKeeper 3.5.10

Ganglia 3.7.2 Hadoop 3.3.6 JupyterEnterpriseGateway 2.6.0 MXNet 1.9.1 Pig 0.17.0 Sqoop 1.4.7 Trino 426

HBase 2.4.17 Hive 3.1.3 JupyterHub 1.5.0 Oozie 5.2.1 Presto 0.283 TensorFlow 2.11.0 Zeppelin 0.10.1

- b. Wählen Sie unter Option Cluster scaling and provisioning (Cluster-Skalierung und -Bereitstellung) Use EMR-managed scaling (EMR-verwaltete Skalierung verwenden) aus.

Cluster scaling and provisioning [Info](#)

Set up scaling and provisioning configurations for the core and task node groups for your cluster.

Choose an option

Set cluster size manually
Use this option if you know your workload patterns in advance.

Use EMR-managed scaling
Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.

Use custom automatic scaling
To programmatically scale core and task nodes, create custom automatic scaling policies.

2. Geben Sie im Abschnitt **Software settings** (Softwareeinstellungen) die folgende Konfiguration ein, um Flink Autoscaler zu aktivieren. Stellen Sie für Testszzenarien das Entscheidungsintervall, das Metrikfensterintervall und das Stabilisierungsintervall auf einen niedrigeren Wert ein, sodass der Auftrag sofort eine Skalierungsentscheidung trifft, um die Überprüfung zu vereinfachen.

```
[
  {
    "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. Wählen oder konfigurieren Sie alle anderen Einstellungen nach Ihren Wünschen und erstellen Sie den Flink-Autoscaler-fähigen Cluster.

Autoscaler-Konfigurationen

In diesem Abschnitt werden die meisten Konfigurationen behandelt, die Sie je nach Ihren spezifischen Anforderungen ändern können.

Note

Bei zeitbasierten Konfigurationen wie `time-`, `interval-` und `window-`Einstellungen ist die Standardeinheit, wenn keine Einheit angegeben ist, Millisekunden. Ein Wert von 30 ohne Suffix entspricht also 30 Millisekunden. Geben Sie für andere Zeiteinheiten das entsprechende Suffix `s` für Sekunden, `m` für Minuten oder `h` für Stunden an.

Themen

- [Autoscaler-Loop-Konfigurationen](#)
- [Aggregation von Metriken und Verlaufskonfigurationen](#)
- [Konfigurationen auf Ebene des Auftrags Scheitelpunkts](#)
- [Konfigurationen für die Backlog-Verarbeitung](#)
- [Vorgangskonfigurationen skalieren](#)

Autoscaler-Loop-Konfigurationen

Autoscaler ruft die Metriken auf Ebene des Auftrags Scheitelpunkts alle paar konfigurierbare Zeitintervalle ab, wandelt sie in umsetzbare Maßeinheiten um, schätzt die Parallelität neuer Auftrags Scheitelpunkte und empfiehlt sie dem Auftrags-Scheduler. Metriken werden erst nach dem Neustart des Auftrags und dem Cluster-Stabilisierungsintervall erfasst.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.enabled</code>	<code>false</code>	Aktivieren Sie die automatische Skalierung auf Ihrem Flink-Cluster.	<code>true</code> , <code>false</code>
<code>job.autoscaler.decision.interval</code>	<code>60s</code>	Autoscaler-Entscheidungsintervall.	30 (Standardeinheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.restart.time</code>	<code>3m</code>	Die erwartete Neustartzeit wird verwendet, bis der Bediener	30 (Standardeinheit ist

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
		sie anhand des Verlaufs zuverlässig ermitteln kann.	Millisekunden), 5m, 1h
<code>job.autoscaler.stabilization.interval</code>	300s	Stabilisierungszeitraum, in dem keine neue Skalierung durchgeführt wird.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.debug.logs.interval</code>	300s	Intervall für Autoscaler-Debug-Protokolle.	30 (Standard einheit ist Millisekunden), 5m, 1h

Aggregation von Metriken und Verlaufskonfigurationen

Autoscaler ruft die Metriken ab, aggregiert sie über ein zeitbasiertes gleitendes Fenster und wertet sie zu Skalierungsentscheidungen aus. Der Verlauf der Skalierungsentscheidungen für jeden Auftrags Scheitelpunkt wird verwendet, um die neue Parallelität abzuschätzen. Diese haben sowohl ein altersabhängiges Verfallsdatum als auch eine historische Größe (mindestens 1).

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.metrics.window</code>	600s	Scaling metrics aggregation window size.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.history.max.count</code>	3	Maximale Anzahl vergangener Skalierungsentscheidungen, die pro Scheitelpunkt beibehalten werden sollen.	1 auf Integer.MAX_VALUE
<code>job.autoscaler.history.max.age</code>	24h	Mindestanzahl vergangener Skalierungsentscheidungen, die pro Scheitelpunkt beibehalten werden sollen.	30 (Standard einheit ist

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
			Millisekunden), 5m, 1h

Konfigurationen auf Ebene des Auftragsscheitelpunkts

Die Parallelität jedes Auftragsscheitelpunkts wird auf der Grundlage der Zielauslastung geändert und durch die Min./Max.-Parallelitätsgrenzen begrenzt. Es wird nicht empfohlen, die Zielauslastung auf nahezu 100 % festzulegen (d. h. den Wert 1), und die Nutzungsgrenze dient als Puffer, um die zwischenzeitlichen Lastschwankungen zu bewältigen.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.target.utilization</code>	0.7	Nutzung des Zielscheitelpunkts.	0 - 1
<code>job.autoscaler.target.utilization.boundary</code>	0.4	Nutzung des Zielscheitelpunkts. Die Skalierung wird nicht durchgeführt, wenn die aktuelle Verarbeitungsrate innerhalb von $[\text{target_rate} / (\text{target_utilization} - \text{boundary})$ und $(\text{target_rate} / (\text{target_utilization} + \text{boundary}))$ liegt	0 - 1
<code>job.autoscaler.vertex.min-parallelism</code>	1	Die Mindestparallelität, die der Autoscaler verwenden kann.	0 - 200
<code>job.autoscaler.vertex.max-parallelism</code>	200	Die maximale Parallelität, die der Autoscaler verwenden kann. Der Autoscaler ignoriert dieses Limit, wenn es höher ist als die maximale Parallelität,	0 - 200

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
		die in der Flink-Konfiguration oder direkt für jeden Bediener konfiguriert wurde.	

Konfigurationen für die Backlog-Verarbeitung

Der Auftrags Scheitelpunkt benötigt zusätzliche Ressourcen, um die ausstehenden Ereignisse oder Rückstände zu bewältigen, die sich während des Skalierungsvorgangs ansammeln.

Dies wird auch als die `catch-up`-Dauer bezeichnet. Wenn die Zeit für die Verarbeitung des Backlogs den konfigurierten `lag -threshold`-Wert überschreitet, steigt die Zielauslastung des Auftrags Scheitelpunkts auf den Höchstwert. Auf diese Weise können unnötige Skalierungsvorgänge während der Backlog-Verarbeitung vermieden werden.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.backlog-processing.lag-threshold</code>	5m	Verzögerungsschwellenwert, der unnötige Skalierungen verhindert und gleichzeitig die ausstehenden Nachrichten entfernt, die für die Verzögerung verantwortlich sind.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.catch-up.duration</code>	15m	Die Zieldauer für die vollständige Backlog-Verarbeitung nach einem Skalierungsvorgang. Auf 0 setzen, um die Backlog-basierte Skalierung zu deaktivieren.	30 (Standard einheit ist Millisekunden), 5m, 1h

Vorgangskonfigurationen skalieren

Autoscaler führt die Herunterskalierung innerhalb einer Kulanzzzeit nicht unmittelbar nach einer Hochskalierung durch. Dadurch wird ein unnötiger Zyklus von Hoch- und

Herunterskalierungsvorgängen, die durch vorübergehende Lastschwankungen verursacht werden, vermieden.

Wir können das Verhältnis der Herunterskalierung verwenden, um die Parallelität schrittweise zu verringern und Ressourcen freizusetzen, um vorübergehende Lastspitzen auszugleichen. Das trägt auch dazu bei, unnötige kleinere Hochskalierungsvorgänge nach einer umfangreichen Herunterskalierung zu vermeiden.

Wir können einen ineffektiven Skalierungsvorgang auf der Grundlage vergangener Entscheidungen über die Skalierung von Auftragsscheitelpunkten erkennen, um weitere Änderungen der Parallelität zu verhindern.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.scale-up.grace-period</code>	1h	Dauer, in der ein Scheitelpunkt nicht herunterskaliert werden darf, nachdem er hochskaliert wurde.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.scale-down.max-factor</code>	0.6	Maximaler Herunterskalierungsfaktor. Ein Wert von 1 bedeutet, dass es keine Begrenzung der Herunterskalierung gibt. 0.6 bedeutet, dass der Auftrag nur mit 60 % der ursprünglichen Parallelität herunterskaliert werden kann.	0 - 1
<code>job.autoscaler.scale-up.max-factor</code>	100000.	Maximales Verhältnis der Hochskalierung. Der Wert 2.0 bedeutet, dass der Auftrag nur mit 200 % der aktuellen Parallelität hochskaliert werden kann.	0 - Integer.MAX_VALUE
<code>job.autoscaler.scaling.effectiveness.detection.enabled</code>	false	Ob ineffektive Skalierungsvorgänge erkannt werden sollen und ob der Autoscale	true, false

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
		r weitere Hochskalierungen blockieren kann.	

Optimieren der Neustartzeiten von Aufträgen für die Aufgabenwiederherstellung und -skalierung

Wenn eine Aufgabe fehlschlägt oder wenn ein Skalierungsvorgang stattfindet, versucht Flink, die Aufgabe vom letzten abgeschlossenen Prüfpunkt aus erneut auszuführen. Die Ausführung des Neustartvorgangs kann eine Minute oder länger dauern, abhängig von der Größe des Prüfpunktzustands und der Anzahl der parallelen Aufgaben. Während des Neustarts können sich Backlog-Aufgaben für den Auftrag ansammeln. Es gibt jedoch einige Möglichkeiten, wie Flink die Geschwindigkeit der Wiederherstellung und des Neustarts von Ausführungsdiagrammen optimiert, um die Auftragsstabilität zu verbessern.

Auf dieser Seite werden einige der Möglichkeiten beschrieben, mit denen Amazon EMR Flink die Zeit für den Neustart des Auftrags während der Aufgabenwiederherstellung oder -skalierung verbessern kann.

Themen

- [Aufgabenlokale Wiederherstellung](#)
- [Generischer protokollbasierter inkrementeller Prüfpunkt](#)
- [Differenzierte Wiederherstellung](#)
- [Kombinierter Neustartmechanismus im adaptiven Scheduler](#)

Aufgabenlokale Wiederherstellung

Note

Aufgabenlokale Wiederherstellung wird mit Amazon EMR 6.0.0 und höher unterstützt.

Mit Flink-Prüfpunkten erstellt jede Aufgabe einen Snapshot ihres Status, den Flink in verteilte Speicher wie Amazon S3 schreibt. Im Falle einer Wiederherstellung stellen die Aufgaben ihren Status

aus dem verteilten Speicher wieder her. Der verteilte Speicher bietet Fehlertoleranz und kann den Status während der Neuskalierung neu verteilen, da er für alle Knoten zugänglich ist.

Ein verteilter Remote-Speicher hat jedoch auch einen Nachteil: Alle Aufgaben müssen ihren Status von einem entfernten Standort aus über das Netzwerk lesen. Dies kann bei der Aufgabenwiederherstellung oder bei Skalierungsvorgängen zu langen Wiederherstellungszeiten für große Zustände führen.

Dieses Problem der langen Wiederherstellungszeit wird durch eine aufgabenlokale Wiederherstellung gelöst. Aufgaben schreiben ihren Status am Prüfpunkt in einen sekundären Speicher, der sich lokal zur Aufgabe befindet, z. B. auf eine lokale Festplatte. Sie speichern ihren Status auch im Primärspeicher oder in unserem Fall in Amazon S3. Während der Wiederherstellung plant der Scheduler die Aufgaben in demselben Task-Manager, in dem die Aufgaben zuvor ausgeführt wurden, sodass sie aus dem lokalen Statusspeicher wiederhergestellt werden können, anstatt sie aus dem Remote-Statusspeicher zu lesen. Weitere Informationen finden Sie unter [Aufgabenlokale Wiederherstellung](#) in der Apache-Flink-Dokumentation.

Unsere Benchmark-Tests mit Beispielaufträgen haben gezeigt, dass die Wiederherstellungszeit bei aktivierter aufgabenlokaler Wiederherstellung von Minuten auf wenige Sekunden reduziert wurde.

Um die aufgabenlokale Wiederherstellung zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer `flink-conf.yaml`-Datei fest. Geben Sie den Wert für das Prüfpunkt-Intervall in Millisekunden an.

```
state.backend.local-recovery: true
state.backend: hasmap or rocksdb
state.checkpoints.dir: s3://storage-location-bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

Generischer protokollbasierter inkrementeller Prüfpunkt

Note

Generische protokollbasierte inkrementelle Prüfpunkte werden mit Amazon EMR 6.10.0 und höher unterstützt.

Generische protokollbasierte inkrementelle Prüfpunkte wurden in Flink 1.16 hinzugefügt, um die Geschwindigkeit von Prüfpunkten zu verbessern. Ein schnelleres Prüfpunktintervall führt

häufig zu einer Reduzierung des Wiederherstellungsaufwands, da weniger Ereignisse nach der Wiederherstellung erneut verarbeitet werden müssen. Weitere Informationen finden Sie im Apache-Flink-Blog unter [Verbesserung der Geschwindigkeit und Stabilität von Prüfpunkten mit generischen protokollbasierten inkrementellen Prüfpunkten](#).

Unsere Benchmark-Tests haben anhand von Beispielaufträgen gezeigt, dass sich die Prüfpunktzeit mit dem generischen protokollbasierten inkrementellen Prüfpunkt von Minuten auf wenige Sekunden reduziert hat.

Um generische protokollbasierte inkrementelle Prüfpunkte zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer Datei `flink-conf.yaml` fest. Geben Sie den Wert für das Prüfpunktintervall in Millisekunden an.

```
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
```

Differenzierte Wiederherstellung

Note

Eine differenzierte Wiederherstellungsunterstützung für den Standard-Scheduler ist mit Amazon EMR 6.0.0 und höher verfügbar. Eine differenzierte Wiederherstellungsunterstützung im adaptiven Scheduler ist mit Amazon EMR 6.15.0 und höher verfügbar.

Wenn eine Aufgabe während der Ausführung fehlschlägt, setzt Flink das gesamte Ausführungsdiagramm zurück und löst eine vollständige Neuausführung ab dem letzten abgeschlossenen Prüfpunkt aus. Das ist teurer, als nur die fehlgeschlagenen Aufgaben erneut auszuführen. Bei einer differenzierten Wiederherstellung wird nur die mit der Pipeline verbundene Komponente der fehlgeschlagenen Aufgabe neu gestartet. Im folgenden Beispiel hat das Auftragsdiagramm 5 Scheitelpunkte (A bis E). Alle Verbindungen zwischen den Scheitelpunkten werden punktweise in Pipelines verlegt, und der Wert `parallelism.default` für den Auftrag ist auf 2 eingestellt.

```
A # B # C # D # E
```

In diesem Beispiel werden insgesamt 10 Aufgaben ausgeführt. Die erste Pipeline (a1 bis e1) wird in einem TaskManager (TM1) ausgeführt, während die zweite Pipeline (a2 bis e2) in einem anderen TaskManager (TM2) ausgeführt wird.

```
a1 # b1 # c1 # d1 # e1  
a2 # b2 # c2 # d2 # e2
```

Es gibt zwei Komponenten, die über eine Pipeline miteinander verbunden sind: a1 # e1 und a2 # e2. Wenn entweder TM1 oder TM2 fehlschlägt, wirkt sich der Fehler nur auf die 5 Aufgaben in der Pipeline aus, in denen der TaskManager ausgeführt wurde. Bei der Neustartstrategie wird nur die betroffene Pipeline-Komponente gestartet.

Eine differenzierte Wiederherstellung funktioniert nur mit perfekt parallelen Flink-Aufträgen. Sie wird nicht mit `keyBy()`- oder `reDistribute()`-Vorgängen unterstützt. Weitere Informationen finden Sie unter [FLIP-1: Fine Grained Recovery from Task Failures](#) (FLIP-1: Differenzierte Wiederherstellung nach Aufgabenfehlern) im Jira-Projekt Flink Improvement Proposal.

Um die differenzierte Wiederherstellung zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer `flink-conf.yaml`-Datei fest.

```
jobmanager.execution.failover-strategy: region  
restart-strategy: exponential-delay or fixed-delay
```

Kombinierter Neustartmechanismus im adaptiven Scheduler

Note

Der kombinierte Neustartmechanismus im adaptiven Scheduler wird mit Amazon EMR 6.15.0 und höher unterstützt.

Der adaptive Scheduler kann die Parallelität des Auftrags auf der Grundlage der verfügbaren Slots anpassen. Er reduziert automatisch die Parallelität, wenn nicht genügend Slots für die konfigurierte Auftragsparallelität verfügbar sind. Wenn neue Slots verfügbar werden, wird der Auftrag wieder auf die konfigurierte Auftragsparallelität hochskaliert. Ein adaptiver Scheduler vermeidet Ausfallzeiten

beim Auftrag, wenn nicht genügend Ressourcen verfügbar sind. Dies ist der unterstützte Scheduler für Flink Autoscaler. Aus diesen Gründen empfehlen wir den adaptiven Scheduler mit Amazon EMR Flink. Adaptive Scheduler können jedoch innerhalb kurzer Zeit mehrere Neustarts durchführen, und zwar einen Neustart für jede neu hinzugefügte Ressource. Dies könnte zu einem Leistungsabfall des Auftrags führen.

Mit Amazon EMR 6.15.0 und höher verfügt Flink über einen kombinierten Neustartmechanismus im adaptiven Scheduler, der ein Neustartfenster öffnet, wenn die erste Ressource hinzugefügt wird, und dann bis zum konfigurierten Fensterintervall von 1 Minute wartet. Er führt einen einzigen Neustart durch, wenn genügend Ressourcen zur Verfügung stehen, um den Auftrag mit konfigurierter Parallelität auszuführen, oder wenn das Intervall abgelaufen ist.

Unsere Benchmark-Tests haben anhand von Beispielaufträgen gezeigt, dass dieses Feature 10 % mehr Datensätze verarbeitet als das Standardverhalten, wenn Sie den adaptiven Scheduler und Flink Autoscaler verwenden.

Um den kombinierten Neustartmechanismus zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer Datei `flink-conf.yaml` fest.

```
jobmanager.adaptive-scheduler.combined-restart.enabled: true
jobmanager.adaptive-scheduler.combined-restart.window-interval: 1m
```

Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR

Einführung

Amazon-EMR-Versionen 6.10.0 und höher unterstützen die [Apache Zeppelin](#)-Integration mit Apache Flink. Sie können Flink-Aufträge interaktiv über Zeppelin-Notebooks einreichen. Mit dem Flink-Interpreter können Sie Flink-Abfragen ausführen, Flink-Streaming- und Batch-Aufträge definieren und die Ausgabe in Zeppelin-Notebooks visualisieren. Der Flink-Interpreter basiert auf der Flink-REST-API. Auf diese Weise können Sie von der Zeppelin-Umgebung aus auf Flink-Aufträge zugreifen und diese bearbeiten, um eine Datenverarbeitung und -analyse in Echtzeit durchzuführen.

In Flink Interpreter gibt es vier Unterinterpreter. Sie dienen unterschiedlichen Zwecken, befinden sich aber alle in der JVM und teilen sich dieselben vorkonfigurierten Einstiegspunkte zu Flink (`ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`). Die Interpreter sind wie folgt:

- `%flink` – Erzeugt `ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment` und stellt eine Scala-Umgebung bereit
- `%flink.pyflink` – Stellt eine Python-Umgebung bereit
- `%flink.ssql` – Stellt eine Streaming-SQL-Umgebung bereit
- `%flink.bsql` – Stellt eine Batch-SQL-Umgebung bereit

Voraussetzungen

- Die Zeppelin-Integration mit Flink wird für Cluster unterstützt, die mit Amazon EMR 6.10.0 und höher erstellt wurden.
- Um Webschnittstellen, die auf EMR-Clustern gehostet werden, wie für diese Schritte erforderlich, anzuzeigen, müssen Sie einen SSH-Tunnel konfigurieren, der eingehenden Zugriff ermöglicht. Weitere Informationen finden Sie unter [Konfigurieren von Proxy-Einstellungen, um auf dem Primärknoten gehostete Websites anzeigen zu lassen](#).

Zeppelin-Flink auf einem EMR-Cluster konfigurieren

Gehen Sie wie folgt vor, um Apache Flink auf Apache Zeppelin für die Ausführung auf einem EMR-Cluster zu konfigurieren:


1. Erstellen Sie einen neuen Cluster von der Amazon-EMR-Konsole aus. Wählen Sie `emr-6.10.0` oder höher für die Amazon-EMR-Version aus. Wählen Sie dann, ob Sie Ihr Anwendungspaket mit der Option Benutzerdefiniert anpassen möchten. Nehmen Sie mindestens Flink, Hadoop und Zeppelin in Ihr Paket auf.


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


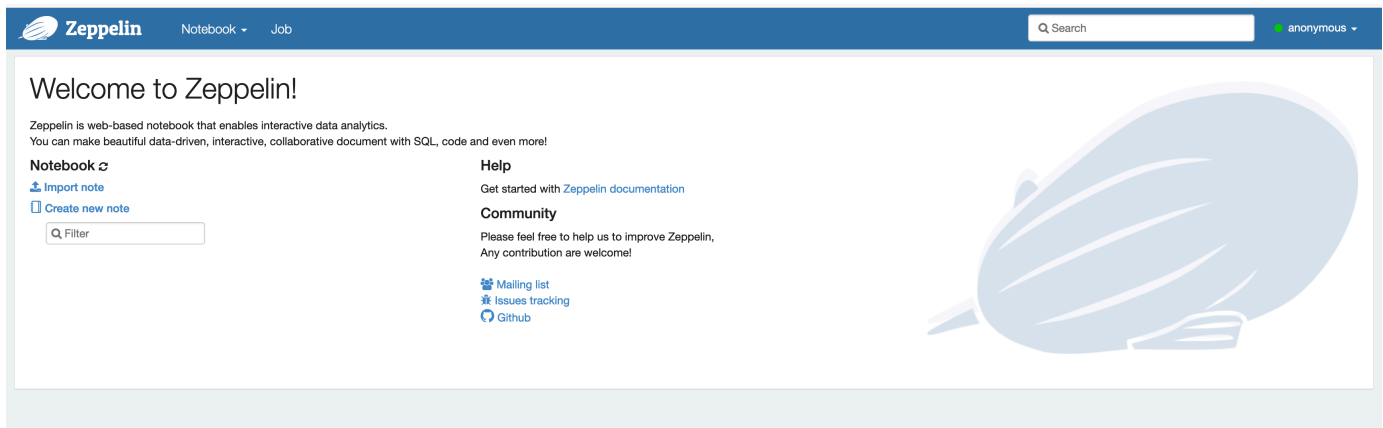
Custom


▼ Customize your application bundle

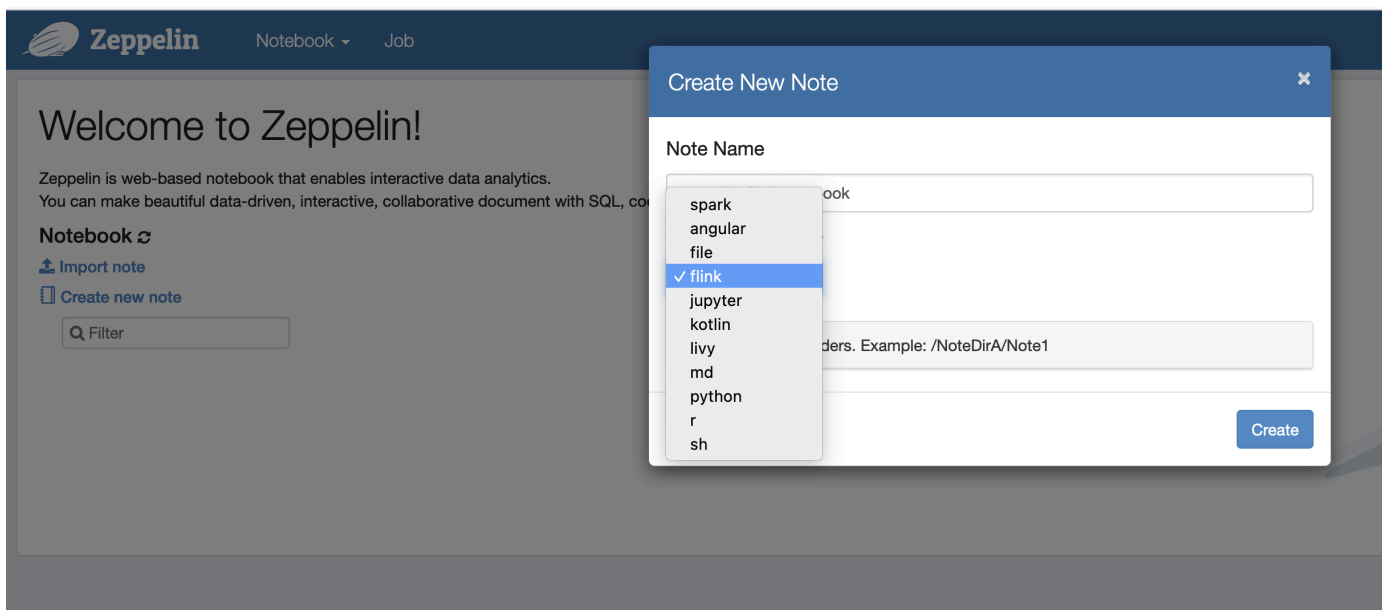
Applications included in bundle

<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. Erstellen Sie den Rest Ihres Clusters mit den Einstellungen, die Sie bevorzugen.
3. Sobald Ihr Cluster läuft, wählen Sie den Cluster in der Konsole aus, um seine Details anzuzeigen, und öffnen Sie die Registerkarte Anwendungen. Wählen Sie Zeppelin im Bereich Benutzeroberflächen für Anwendungen aus, um die Zeppelin-Weboberfläche zu öffnen. Stellen Sie sicher, dass Sie den Zugriff auf die Zeppelin-Weboberfläche mit einem SSH-Tunnel zum Primärknoten und einer Proxyverbindung eingerichtet haben, wie in [Voraussetzungen](#) beschrieben.



4. Jetzt können Sie eine neue Notiz in einem Zeppelin-Notebook mit Flink als Standardinterpreter erstellen.



5. In den folgenden Codebeispielen wird veranschaulicht, wie Flink-Jobs von einem Zeppelin-Notebook aus ausgeführt werden.

Führen Sie Flink-Aufträge mit Zeppelin-Flink auf einem EMR-Cluster aus

- Beispiel 1, Flink Scala

a) Batch- WordCount Beispiel (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) Streaming- WordCount Beispiel (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()
```

- Beispiel 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. The top part displays the SQL code used for the job. Below the code, a table shows the results of the query. The table has two columns: 'id' and 'data'. The data is as follows:

id	data
3	Hello
3	World
3	Hello
3	World
4	ADD
4	ADD
5	LINE
5	LINE

At the bottom of the screenshot, it indicates the job took 20 seconds and was last updated by an anonymous user on May 04, 2023, at 3:07:03 PM.

- Beispiel 3, Pyflink. Beachten Sie, dass Sie Ihre eigene Beispieldatei mit dem Namen `word.txt` in Ihren S3-Bucket hochladen müssen.

```
%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. Wählen Sie FLINK JOB in der Zeppelin-Benutzeroberfläche, um auf die Flink-Web-UI zuzugreifen und diese anzusehen.



Batch WordCount FLINK JOB FINISHED ▶ ⌵ 📖 ⚙️

```

%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()

```

data: org.apache.flink.api.scala.DataSet[String] = org.apache.flink.api.scala.DataSet@22fe7dd5
(flink,1)
(hadoop,1)
(hello,3)
(world,1)

Took 56 sec. Last updated by anonymous at May 04 2023, 2:19:24 PM. (outdated)

2. Wenn Sie FLINK JOB wählen, gelangen Sie zur Flink Web Console in einer anderen Registerkarte Ihres Browsers.

The screenshot shows the Apache Flink Dashboard interface. On the left is a navigation sidebar with options: Overview, Jobs (selected), 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

Versionsverlauf von Flink

In der folgenden Tabelle sind die Version von Flink aufgeführt, die in jeder Release-Version von Amazon enthalten ist EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Flink-Versionsinformationen

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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,

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
		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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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-ya rn-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-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
emr-6.14.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.13.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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-ya rn-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-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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-yarn-timeline-server, flink-client
emr-5.28.1	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.28.0	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
emr-5.21.2	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.1	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.0	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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-resource manager, 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-resource manager, 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-resource manager, flink-client

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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-resource manager, 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-resource manager, 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-resource manager, flink-client

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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

Amazon EMR Release-Etikett	Flink-Version	Mit Flink installierte Komponenten
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-resource manager, flink-client

Versionshinweise zu Flink nach Version

Die vollständigen Versionshinweise finden Sie in den folgenden Abschnitten.

Amazon EMR 7.2.0 — Versionshinweise zu Flink

Typ	Beschreibung
Verbesserung	Support das Hinzufügen von benutzerdefinierten Labels zum Kubernetes-Dienst pro Flink-Job über die Konfiguration. <code>kubernetes.service.labels</code>

Ganglia

Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

Das Open-Source-Projekt Ganglia ist ein skalierbares, verteiltes System zur Überwachung von Clustern und Grids, das zugleich die Auswirkungen auf die Leistung minimiert. Wenn Sie Ganglia in Ihrem Cluster aktivieren, können Sie Berichte erstellen und die Leistung des Clusters als Ganzes betrachten. Ebenso können Sie die Leistung einzelner Knoten-Instances überprüfen. Ganglia ist außerdem zur Aufnahme und Visualisierung von Hadoop- und Spark-Metriken konfiguriert. Weitere Informationen zum Open-Source-Projekt Ganglia finden Sie unter <http://ganglia.info/>.

Wenn Sie die Ganglia-Web-Benutzeroberfläche in einem Browser anzeigen, sehen Sie eine Übersicht über die Cluster-Leistung mit Diagrammen, die Details zur Last, Arbeitsspeicherauslastung, CPU-Auslastung und zum Netzwerkverkehr des Clusters enthalten. Unter den Cluster-Statistiken werden Diagramme für jeden einzelnen Server im Cluster angegeben.

Die folgende Tabelle listet die Version von Ganglia auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Ganglia installiert.

Die Version der Komponenten, die mit Ganglia in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Ganglia-Versioneninformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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-

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
		server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Die folgende Tabelle listet die Version von Ganglia auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Ganglia installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Ganglia installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

Ganglia-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.36.2	Ganglia 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-resource-manager, hadoop-yarn-timeline-server, webserver

Themen

- [Erstellen eines Clusters mit Ganglia](#)
- [Anzeigen von Ganglia-Metriken](#)
- [Hadoop- und Spark-Metriken in Ganglia](#)
- [Versionsverlauf von Ganglia](#)

Erstellen eines Clusters mit Ganglia

Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

So erstellen Sie einen Cluster mit Ganglia über die Konsole

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Cluster erstellen.
3. Wählen Sie unter Software configuration (Softwarekonfiguration) entweder All Applications (Alle Anwendungen), Core Hadoop oder Spark aus.
4. Fahren Sie mit der Erstellung des Clusters mit Konfigurationen nach Bedarf fort.

Um Ganglia zu einem Cluster hinzuzufügen, verwenden Sie AWS CLI

In der AWS CLI können Sie Ganglien zu einem Cluster hinzufügen, indem Sie `create-cluster` den `--applications` Parameter verwenden. Wenn Sie nur Ganglia mit dem Parameter `--applications` angeben, ist Ganglia die einzige installierte Anwendung.

- Geben Sie den folgenden Befehl ein, um Ganglia hinzuzufügen, wenn Sie einen Cluster erstellen. Ersetzen Sie *myKey* durch den Namen Ihres EC2-Schlüsselpaars.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Wenn Sie die Instance-Anzahl ohne den Parameter `--instance-groups` angeben, wird ein einzelner Master-Knoten gestartet. Die verbleibenden Instances werden als Core-Knoten aufgerufen. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

Note

Wenn Sie zuvor nicht die standardmäßige EMR-Servicerolle und das EC2-Instance-Profil erstellt haben, geben Sie `aws emr create-default-roles` ein, um sie zu erstellen, bevor Sie den Unterbefehl `create-cluster` eingeben.

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen finden Sie AWS CLI unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Anzeigen von Ganglia-Metriken

Note

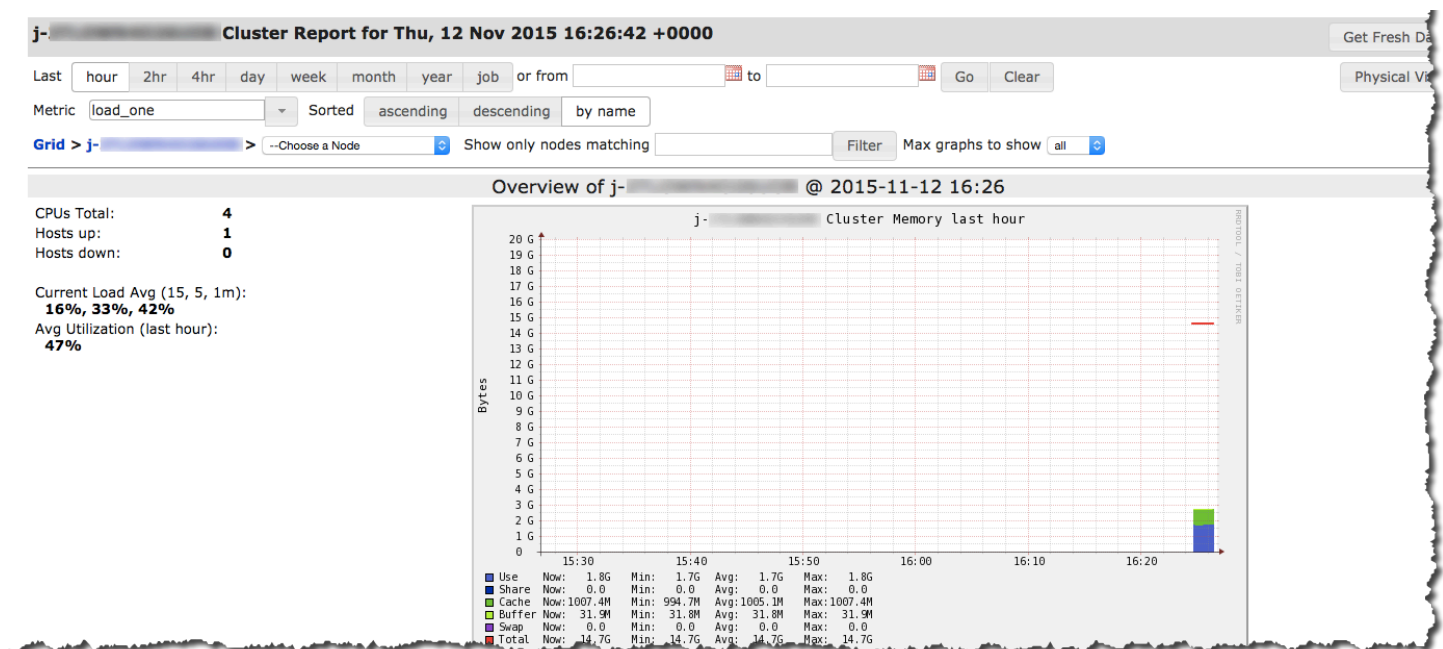
Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

Ganglia bietet eine webbasierte Benutzeroberfläche, die Sie verwenden können, um Metriken Ganglia anzuzeigen. Wenn Sie Ganglia auf Amazon EMR ausführen, wird die Web-Schnittstelle auf dem Hauptknoten ausgeführt und kann über die Port-Weiterleitung (Erstellung eines SSH-Tunnels) angezeigt werden. Weitere Informationen zum Anzeigen von Webbenutzeroberflächen auf finden Sie unter [Auf Amazon-EMR-Clustern gehostete Webschnittstellen anzeigen](#) im Verwaltungshandbuch für Amazon EMR.

So zeigen Sie die Ganglia Weboberfläche an

1. Verwenden Sie SSH, um einen Tunnel zum Master-Knoten und eine sichere Verbindung zu erstellen. Weitere Informationen zum Erstellen eines SSH-Tunnels zum Hauptknoten finden

- Sie unter [Option 2, Teil 1: Einrichten eines SSH-Tunnels zum Hauptknoten mit dynamischer Portweiterleitung](#) im Verwaltungshandbuch für Amazon EMR.
2. Installieren Sie einen Webbrowser mit einem Proxy-Tool, z. B. dem FoxyProxy Plug-in für Firefox, um einen SOCKS-Proxy für Domains des Typs `*ec2*.amazonaws.com*` zu erstellen. Weitere Informationen finden Sie unter [Option 2, Teil 2: Proxy-Einstellungen konfigurieren, um auf dem Hauptknoten gehostete Websites anzeigen zu lassen](#) im Verwaltungshandbuch für Amazon EMR.
 3. Wenn der Proxy gesetzt und die SSH-Verbindung geöffnet ist, können Sie die Ganglia-Benutzeroberfläche anzeigen, indem Sie ein Browserfenster mit `http://master-public-dns-name/ganglia/` öffnen, in dem sich die öffentliche DNS-Adresse des Masterservers im EMR-Cluster `master-public-dns-name` befindet.



Hadoop- und Spark-Metriken in Ganglia

Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

Ganglia meldet Hadoop-Metriken für jede Instance. Die verschiedenen Arten von Metriken sind nach Kategorien mit einem Präfix versehen: verteiltes Dateisystem (dfs.*), Java Virtual Machine (jvm.*), (mapred.*) und Remote Procedure Calls MapReduce (rpc.*).

Auf YARN basierende Ganglia-Metriken wie Spark und Hadoop sind für die EMR-Versionen 4.4.0 und 4.5.0 nicht verfügbar. Verwenden Sie eine höhere Version, um diese Metriken zu verwenden.

Ganglia-Metriken für Spark haben in der Regel Präfixe für YARN-Anwendungs-ID und Spark-DAGScheduler. Die Präfixe haben also folgende Form:

- DAGScheduler.*
- application_XXXXXXXXXX_XXXX.driver.*
- application_XXXXXXXXXX_XXXX.executor.*

Versionsverlauf von Ganglia

Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

In der folgenden Tabelle sind die Version von Ganglia aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Informationen zur Ganglia-Version

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.36.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode,

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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-resource-manager, 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-resource-manager, 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-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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-resource-manager, 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-resource-manager, 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-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
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

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.2.0	3.6.0	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Apache Hadoop

[Apache Hadoop](#) ist ein Open-Source-Java-Software-Framework, das die Verarbeitung von großen Datenmengen mithilfe von Instance-Clustern unterstützt. Hadoop kann auf einer einzigen Instance oder auch auf Tausenden von Instances ausgeführt werden. Hadoop verwendet verschiedene Verarbeitungsmodelle wie MapReduce und Tez, um die Verarbeitung auf mehrere Instanzen zu verteilen, und verwendet auch ein verteiltes Dateisystem, das HDFS zum Speichern von Daten auf mehrere Instanzen aufgerufen wird. Hadoop überwacht den Zustand der Instances im Cluster und kann nach dem Ausfall eines oder mehrerer Knoten eine Wiederherstellung durchführen. Auf diese Weise bietet Hadoop nicht nur erhöhte Datenverarbeitungs- und Speicherkapazitäten, sondern auch eine hohe Verfügbarkeit. Weitere Informationen finden Sie in der [Hadoop-Dokumentation](#).

In der folgenden Tabelle sind die Version von Hadoop aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Hadoop EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hadoop installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Hadoop-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-7.2.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

In der folgenden Tabelle sind die Version von Hadoop aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Hadoop EMR installiert.

Die Version der Komponenten, die mit Hadoop in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Hadoop-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Hadoop-Version	Mit Hadoop installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

In der folgenden Tabelle sind die Version von Hadoop aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Hadoop EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hadoop installiert wurden, finden Sie unter Komponentenversionen der Version 5.36.2.](#)

Hadoop-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.36.2	Hadoop 2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-

EMRAmazon-Freigabeetikett	Hadoop-Version	Mit Hadoop installierte Komponenten
		hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Ab Amazon EMR 5.18.0 können Sie das Amazon EMR Artifact Repository verwenden, um Ihren Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die in bestimmten Amazon-Versionen verfügbar sind. EMR Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Themen

- [Konfigurieren von Hadoop](#)
- [Transparente Verschlüsselung HDFS bei Amazon EMR](#)
- [Erstellen oder Ausführen einer Hadoop-Anwendung](#)
- [Wiederhergestellte Objekte lesen](#)
- [Aktivieren Sie die Erkennung ungleichmäßiger Speicherzugriffe für Container YARN](#)
- [Hadoop-Versionsverlauf](#)

Konfigurieren von Hadoop

Die folgenden Abschnitte enthalten die Standardkonfigurationseinstellungen für Hadoop-Daemons, Tasks und. HDFS

Themen

- [Aufgabenkonfiguration](#)
- [Hadoop-Daemon-Konfigurationseinstellungen](#)
- [HDFSKonfiguration](#)

Aufgabenkonfiguration

Sie können Konfigurationsvariablen festlegen, um die Leistung Ihrer MapReduce Jobs zu optimieren. Dieser Abschnitt enthält die Standardwerte für wichtige Einstellungen. Die Standardwerte variieren je nach EC2 Instanztyp des im Cluster verwendeten Knotens. HBase ist verfügbar, wenn Sie die EMR Amazon-Release-Version 4.6.0 und höher verwenden. Bei HBase der Installation werden unterschiedliche Standardwerte verwendet. Diese Werte werden zusammen mit den anfänglichen Standardwerten bereitgestellt.

Hadoop 2 verwendet zwei Parameter, `mapreduce.map.java.opts` und `mapreduce.reduce.java.opts`, um den Speicher für Map bzw. JVMs Reduce zu konfigurieren. Diese ersetzen die einzelne `mapreduce.map.java.opts`-Konfigurationsoption von früheren Hadoop-Versionen.

Gleichermaßen ersetzt `mapred.job.jvm.num.tasks` `mapred.job.reuse.jvm.num.tasks` in Hadoop 2.7.2 und höher. Amazon EMR legt diesen Wert unabhängig vom EC2 Instance-Typ auf 20 fest. Sie können diese Einstellung mithilfe der `mapred-site`-Konfigurationsklassifizierung überschreiben. Die Einstellung eines Werts von `-1` bedeutet, dass `a` für eine unendliche Anzahl von Aufgaben innerhalb eines einzelnen Jobs wiederverwendet werden JVM kann, und ein Wert von `1` gibt an, dass für jede Aufgabe eine neue Aufgabe erzeugt JVM wird.

Beispiel: Um den Wert von `mapred.job.jvm.num.tasks` auf `-1` festzulegen, können Sie eine Datei mit dem folgenden Inhalt erstellen:

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.job.jvm.num.tasks": "-1"
    }
  }
]
```

Wenn Sie den `create-cluster` Befehl oder den `modify-instance-groups` Befehl aus dem verwenden AWS CLI, können Sie auf die JSON Konfigurationsdatei verweisen. Im folgenden Beispiel wird die Konfigurationsdatei unter `myConfig.json` gespeichert und in Amazon S3 aufbewahrt.

Note

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
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
```

Sie können die unten aufgeführten Standardwerte mithilfe der `mapred-site` Konfigurationsklassifizierung auf dieselbe Weise ändern und mehrere Werte und mehrere Konfigurationsklassifizierungen mithilfe einer einzigen JSON-Datei festlegen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Dazu verwenden Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Standardwerte für die Aufgabenkonfigurationseinstellungen

Instance-Typen

- [C1-Instances](#)
- [C3-Instances](#)
- [C4-Instances](#)
- [C5-Instances](#)
- [C5a-Instances](#)
- [C5ad-Instances](#)
- [C5d-Instances](#)
- [C5n-Instances](#)

- [C6a-Instances](#)
- [C6g-Instances](#)
- [C6gd-Instances](#)
- [C6gn-Instances](#)
- [C6i-Instances](#)
- [C6id-Instances](#)
- [C6in-Instances](#)
- [c7a-Instances](#)
- [C7g-Instances](#)
- [C7gd-Instances](#)
- [C7gn-Instances](#)
- [C7i-Instances](#)
- [D2-Instances](#)
- [D3-Instances](#)
- [D3en-Instances](#)
- [G3-Instances](#)
- [G3s-Instances](#)
- [G4dn-Instances](#)
- [G5-Instances](#)
- [H1-Instances](#)
- [I2-Instances](#)
- [I3-Instances](#)
- [I3en-Instances](#)
- [I4g-Instances](#)
- [I4i-Instances](#)
- [I4gn-Instances](#)
- [I4gen-Instances](#)
- [M1-Instances](#)

- [M2-Instances](#)
- [M3-Instances](#)
- [M4-Instances](#)
- [M5-Instances](#)
- [M5a-Instances](#)
- [M5ad-Instances](#)
- [M5d-Instances](#)
- [M5dn-Instances](#)
- [M5n-Instances](#)
- [M5zn-Instances](#)
- [M6a-Instances](#)
- [M6g-Instances](#)
- [M6gd-Instances](#)
- [M6i-Instances](#)
- [M6id-Instances](#)
- [M6idn-Instances](#)
- [M6in-Instances](#)
- [M7a-Instances](#)
- [M7g-Instances](#)
- [M7gd-Instances](#)
- [M7i-Instances](#)
- [M7i-flex-Instances](#)
- [P2-Instances](#)
- [P3-Instances](#)
- [P5-Instances](#)
- [R3-Instances](#)
- [R4-Instances](#)
- [R5-Instances](#)
- [R5a-Instances](#)

- [R5ad-Instances](#)
- [r5b-Instances](#)
- [R5d-Instances](#)
- [R5dn-Instances](#)
- [R5n-Instances](#)
- [R6a-Instances](#)
- [R6g-Instances](#)
- [R6gd-Instances](#)
- [R6i-Instances](#)
- [R6id-Instances](#)
- [R6idn-Instances](#)
- [R6in-Instances](#)
- [R7a-Instances](#)
- [R7g-Instances](#)
- [R7gd-Instances](#)
- [r7i-Instances](#)
- [R7iz-Instances](#)
- [X1-Instances](#)
- [X1e-Instances](#)
- [X2gd-Instances](#)
- [x2idn-Instances](#)
- [X2iedn-Instances](#)
- [z1d-Instances](#)

C1-Instances

c1.medium

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.map.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>

Konfigurationsoption	Standardwert	Mit HBase installiert
<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
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource.memory-mb</code>	1024	512

c1.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.nodemanager.resource.memory-mb	5120	2560

C3-Instances

c3.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

c3.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

c3.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.nodemanager.resource.memory-mb	23040	11520

c3.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1331m	-Xmx1331m
mapreduce.java.opts	-Xmx2662m	-Xmx2662m
mapreduce.map.memory.mb	1664	1664
mapreduce.reduce.memory.mb	3328	3328
yarn.app.mapreduce.am.resource.mb	3328	3328
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	53248	26624
yarn.nodemanager.resource.memory-mb	53248	26624

C4-Instances

c4.large

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx717m	-Xmx717m
mapreduce.java.opts	-Xmx1434m	-Xmx1434m

Konfigurationsoption	Standardwert	Mit HBase installiert
<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
<code>yarn.nodemanager.resource.memory-mb</code>	1792	896

c4.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.nodemanager.resource.memory-mb	5632	2816

c4.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

c4.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

c4.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

C5-Instances

c5.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

c5.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

c5.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

c5.9xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	65536	32768
yarn.nodemanager.resource.memory-mb	65536	32768

c5.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

c5.18xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c5.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

C5a-Instances

c5a.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c5a.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

c5a.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c5a.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

C5ad-Instances

c5ad.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

c5ad.4xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c5ad.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c5ad.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c5ad.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

C5d-Instances

c5d.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

c5d.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource-memory-mb	12288	6144

c5d.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	65536	32768
yarn.nodemanager.resource.memory-mb	65536	32768

c5d.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiert HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx1547m</code>	<code>-Xmx1547m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3094m</code>	<code>-Xmx3094m</code>
<code>mapreduce.map.memory.mb</code>	1934	1934
<code>mapreduce.reduce.memory.mb</code>	3868	3868
<code>yarn.app.mapreduce.am.resource.mb</code>	3868	3868
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	139264	30960
<code>yarn.nodemanager.resource-memory-mb</code>	139264	30960

c5d.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

C5n-Instances

c5n.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.map.java.opts</code>	<code>-Xmx1613m</code>	<code>-Xmx1613m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3226m</code>	<code>-Xmx3226m</code>
<code>mapreduce.map.memory.mb</code>	2016	2016
<code>mapreduce.reduce.memory.mb</code>	4032	4032
<code>yarn.app.mapreduce.am.resource.mb</code>	4032	4032
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	16128	8064
<code>yarn.nodemanager.resource-memory-mb</code>	16128	8064

c5n.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	34816	17408
yarn.nodemanager.resource.memory-mb	34816	17408

c5n.9xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx2002m	-Xmx2002m
mapreduce.java.opts	-Xmx4004m	-Xmx4004m
mapreduce.map.memory.mb	2503	2503
mapreduce.reduce.memory.mb	5006	5006
yarn.app.mapreduce.am.resource.mb	5006	5006

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx2094m</code>	<code>-Xmx2094m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4188m</code>	<code>-Xmx4188m</code>
<code>mapreduce.map.memory.mb</code>	2617	2617
<code>mapreduce.reduce.memory.mb</code>	5234	5234
<code>yarn.app.mapreduce.am.resource.mb</code>	5234	5234
<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

C6a-Instances

c6a.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6a.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6a.4xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6a.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6a.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c6a.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource-memory-mb	241664	30208

c6a.48xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

C6g-Instances

c6g.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6g.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

c6g.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6g.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6g.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

C6gd-Instances

c6gd.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c6gd.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

c6gd.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6gd.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

C6gn-Instances

c6gn.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c6gn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

c6gn.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6gn.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

C6i-Instances

c6i.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c6i.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

c6i.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6i.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource-memory-mb	85504	32074

c6i.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c6i.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

C6id-Instances

c6id.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c6id.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

c6id.4xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6id.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6id.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

c6id.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

C6in-Instances

c6in.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource-memory-mb	5632	2816

c6in.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6in.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6in.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c6in.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c6in.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7a-Instances

c7a.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7a.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

c7a.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c7a.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

c7a.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7a.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

c7a.48xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

C7g-Instances

c7g.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7g.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c7g.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7g.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

C7gd-Instances

c7gd.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7gd.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

c7gd.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c7gd.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7gd.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

C7gn-Instances

c7gn.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7gn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource-memory-mb	11712	5856

c7gn.4xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c7gn.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c7gn.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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-Instances

c7i.x groß

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7i.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

c7i.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

c7i.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource-memory-mb	85504	32074

c7i.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

c7i.48xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

D2-Instances

d2.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

d2.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

d2.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

d2.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx2417m	-Xmx2417m
mapreduce.java.opts	-Xmx4834m	-Xmx4834m
mapreduce.map.memory.mb	3021	3021
mapreduce.reduce.memory.mb	6042	6042
yarn.app.mapreduce.am.resource.mb	6042	6042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30194
yarn.nodemanager.resource.memory-mb	241664	30194

D3-Instances

d3.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

d3.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

d3.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

d3.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

D3en-Instances

d3en.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

d3en.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

d3en.6xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	28496
yarn.nodemanager.resource-memory-mb	85504	28496

d3en.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

G3-Instances

g3.4xgroß

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

g3.8xgroß

Konfigurationsoption	Standardwert	Mit installiert HBase
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

g3.16xgroß

Konfigurationsoption	Standardwert	Mit installiert HBase
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

G3s-Instances

g3s.xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

G4dn-Instances

g4dn.xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

g4dn.2xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

g4dn.4xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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.8xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

g4dn.12xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

g4dn.16xgroß

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

G5-Instances

g5.xgroß

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

g5.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

g5.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

g5.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

g5.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

H1-Instances

h1.2xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource-memory-mb	24576	12288

h1.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

h1.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

h1.16xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

I2-Instances

i2.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

i2.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

i2.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

I3-Instances

i3.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

i3.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

i3.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

i3.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

i3.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

I3en-Instances

i3en.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

i3en.3xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx6007m	-Xmx6007m
mapreduce.java.opts	-Xmx12014m	-Xmx12014m
mapreduce.map.memory.mb	7509	7509
mapreduce.reduce.memory.mb	15018	15018
yarn.app.mapreduce.am.resource.mb	15018	15018

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource-memory-mb	90112	30040

i3en.6xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

i3en.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

i3en.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

I4g-Instances

i4g.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

i4g.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

i4g.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

i4g.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

i4g.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

I4i-Instances

i4i.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource-memory-mb	23424	11712

i4i.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

i4i.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

i4i.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

i4i.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

i4i.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

i4i.24xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource-memory-mb	741376	30860

i4i.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource-memory-mb	991232	30976

Im4gn-Instances

im4gn.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

im4gn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource-memory-mb	23424	11712

im4gn.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

im4gn.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

Is4gen-Instances

is4gen.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx3514m	-Xmx3514m
mapreduce.java.opts	-Xmx7028m	-Xmx7028m
mapreduce.map.memory.mb	4393	4393
mapreduce.reduce.memory.mb	8786	8786
yarn.app.mapreduce.am.resource.mb	8786	8786
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	17572,12	8786,06
yarn.nodemanager.resource.memory-mb	17572,12	8786,06

is4gen.2xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx3866m	-Xmx3866m
mapreduce.java.opts	-Xmx7732m	-Xmx7732m
mapreduce.map.memory.mb	4832	4832
mapreduce.reduce.memory.mb	9664	9664
yarn.app.mapreduce.am.resource.mb	9664	9664
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

is4gen.4xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx4275m	-Xmx4275m
mapreduce.java.opts	-Xmx8550m	-Xmx8550m
mapreduce.map.memory.mb	5344	5344
mapreduce.reduce.memory.mb	10688	10688
yarn.app.mapreduce.am.resource.mb	10688	10688

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32064
yarn.nodemanager.resource.memory-mb	85504	32064

is4gen.8xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

M1-Instances

m1.small

Konfigurationsoption	Standardwert	Mit installiert HBase
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
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	512	512
yarn.nodemanager.resource.memory-mb	1024	512

m1.medium

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx768m	-Xmx768m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	1024	1024
yarn.app.mapreduce.am.resource.mb	1024	1024

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.map.java.opts</code>	<code>-Xmx512m</code>	<code>-Xmx512m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1024m</code>	<code>-Xmx1024m</code>
<code>mapreduce.map.memory.mb</code>	768	768
<code>mapreduce.reduce.memory.mb</code>	1536	1536
<code>yarn.app.mapreduce.am.resource.mb</code>	1536	1536
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	3072	2560
<code>yarn.nodemanager.resource-memory-mb</code>	5120	2560

m1.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	6144
yarn.nodemanager.resource.memory-mb	12288	6144

M2-Instances

m2.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx864m	-Xmx864m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	1024	1024
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	7168	7168
yarn.nodemanager.resource-memory-mb	14336	7168

m2.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1280m	-Xmx1280m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	2560	2560
yarn.app.mapreduce.am.resource.mb	2560	2560
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	15360
yarn.nodemanager.resource-memory-mb	30720	15360

m2.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1280m	-Xmx1280m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	2560	2560
yarn.app.mapreduce.am.resource.mb	2560	2560
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	30720
yarn.nodemanager.resource.memory-mb	61440	30720

M3-Instances

m3.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource-memory-mb	11520	5760

m3.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource-memory-mb	23040	11520

M4-Instances

m4.large

Konfigurationsoption	Standardwert	Mit HBase installiert
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	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

m4.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource-memory-mb	12288	6144

m4.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

m4.10xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx1557m	-Xmx1557m
mapreduce.java.opts	-Xmx3114m	-Xmx3114m
mapreduce.map.memory.mb	1946	1946
mapreduce.reduce.memory.mb	3892	3892
yarn.app.mapreduce.am.resource.mb	3892	3892

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	155648	31104
yarn.nodemanager.resource-memory-mb	155648	31104

m4.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx1587m	-Xmx1587m
mapreduce.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.map.memory.mb	1984	1984
mapreduce.reduce.memory.mb	3968	3968
yarn.app.mapreduce.am.resource.mb	3968	3968
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource-memory-mb	253952	31744

M5-Instances

m5.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m5.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource-memory-mb	24576	12288

m5.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5.12xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource-memory-mb	188416	31416

m5.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

M5a-Instances

m5a.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m5a.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5a.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource-memory-mb	122880	30720

m5a.12xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

m5a.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

M5ad-Instances

m5ad.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5ad.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m5ad.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m5ad.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

M5d-Instances

m5d.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

m5d.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource-memory-mb	24576	12288

m5d.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5d.12xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource-memory-mb	188416	31416

m5d.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

M5dn-Instances

m5dn.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m5dn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5dn.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

m5dn.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

m5dn.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

M5n-Instances

m5n.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m5n.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

m5n.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m5n.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource-memory-mb	241664	30208

m5n.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

M5zn-Instances

m5zn.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.java.opts	-Xmx4608m	-Xmx4608m
mapreduce.map.memory.mb	2880	2880
mapreduce.reduce.memory.mb	5760	5760
yarn.app.mapreduce.am.resource.mb	5760	5760
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

m5zn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m5zn.3xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	28496
yarn.nodemanager.resource.memory-mb	85504	28496

m5zn.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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-Instances

m6a.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6a.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

m6a.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6a.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6a.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6a.48xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

M6g-Instances

m6g.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6g.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

m6g.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6g.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

M6gd-Instances

m6gd.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6gd.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

m6gd.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6gd.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

M6i-Instances

m6i.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6i.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6i.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

m6i.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6i.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6i.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

M6id-Instances

m6id.x groß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource-memory-mb	11712	5856

m6id.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6id.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6id.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m6id.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

m6id.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

M6idn-Instances

m6idn.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6id.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource-memory-mb	23424	11712

m6idn.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6idn.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m6idn.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6idn.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

M6in-Instances

m6in.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m6in.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m6in.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6n.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m6in.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource-memory-mb	241664	30208

m6in.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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-Instances

m7a.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

m7a.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m7a.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m7a.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

m7a.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

m7a.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m7a.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

M7g-Instances

m7g.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7g.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7g.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m7g.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

m7g.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource-memory-mb	179200	29880

m7g.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

M7gd-Instances

m7gd.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7gd.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

m7gd.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

m7gd.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource-memory-mb	179200	29880

m7gd.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

M7i-Instances

m7i.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

m7i.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

m7i.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

m7i.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

M7i-flex-Instances

m7i-flex.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

m7i-flex.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource-memory-mb	23424	11712

m7i-flex.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

P2-Instances

p2.xgroß

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136
yarn.app.mapreduce.am.resource.mb	27136	27136

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

p2.8xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.map.memory.mb	15360	15360
mapreduce.reduce.memory.mb	30720	30720
yarn.app.mapreduce.am.resource.mb	30720	30720
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

p2.16xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx9267m	-Xmx9267m
mapreduce.java.opts	-Xmx18534m	-Xmx18534m
mapreduce.map.memory.mb	11584	11584
mapreduce.reduce.memory.mb	23168	23168
yarn.app.mapreduce.am.resource.mb	23168	23168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	23168
yarn.nodemanager.resource.memory-mb	741376	23168

P3-Instances

p3.2xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

p3.8xgroß

Konfigurationsoption	Standardwert	Mit HBase installiert
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

p3.16xgroß

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

P5-Instances

p5.48xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx8294m	-Xmx8294m
mapreduce.java.opts	-Xmx16588m	-Xmx16588m
mapreduce.map.memory.mb	10368	10368
mapreduce.reduce.memory.mb	20736	20736
yarn.app.mapreduce.am.resource.mb	20736	20736

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	20736
<code>yarn.nodemanager.resource-memory-mb</code>	1990656	20736

R3-Instances

r3.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

r3.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r3.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

r3.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

R4-Instances

r4.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r4.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

r4.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r4.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r4.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

R5-Instances

r5.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

r5.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

r5.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

r5.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

r5.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

R5a-Instances

r5a.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

r5a.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource-memory-mb	57344	28672

r5a.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

r5a.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

r5a.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

r5a.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

R5ad-Instances

r5ad.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource-memory-mb	23424	11712

r5ad.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r5ad.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5ad.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	499712	31232
yarn.nodemanager.resource.memory-mb	499712	31232

r5ad.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

r5b-Instances

r5b.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5b.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r5b.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiert HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource-memory-mb</code>	241664	30208

r5b.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

r5b.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource-memory-mb	491520	30720

r5b.24xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

R5d-Instances

r5d.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

r5d.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

r5d.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r5d.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

r5d.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource-memory-mb	385024	32100

r5d.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource-memory-mb	516096	32256

r5d.24xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

R5dn-Instances

r5dn.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r5dn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r5dn.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource-memory-mb</code>	366592	30564

r5dn.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

r5dn.24xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

R5n-Instances

r5n.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r5n.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

r5n.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

r5n.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

R6a-Instances

r6a.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6a.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r6a.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource-memory-mb	366592	30564

r6a.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r6a.32xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx6212m</code>	<code>-Xmx6212m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12424m</code>	<code>-Xmx12424m</code>
<code>mapreduce.map.memory.mb</code>	7765	7765
<code>mapreduce.reduce.memory.mb</code>	15530	15530
<code>yarn.app.mapreduce.am.resource.mb</code>	15530	15530
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1490944	31124
<code>yarn.nodemanager.resource-memory-mb</code>	1490944	31124

R6g-Instances

r6g.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6g.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6g.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r6g.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6g.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

R6gd-Instances

r6gd.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6gd.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6gd.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r6gd.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource-memory-mb	366592	30564

r6gd.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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-Instances

r6i.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6i.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

r6i.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r6i.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6i.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r6i.32xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6029m	-Xmx6029m
mapreduce.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.map.memory.mb	7536	7536
mapreduce.reduce.memory.mb	15072	15072
yarn.app.mapreduce.am.resource.mb	15072	15072

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

R6id-Instances

r6id.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6id.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r6id.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6id.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

r6id.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

r6id.32xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

R6idn-Instances

r6idn.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6idn.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r6idn.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6idn.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

r6idn.24xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
<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

Konfigurationsoption	Standardwert	Mit installiertem HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource-memory-mb</code>	991232	30976

R6in-Instances

r6in.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r6in.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource-memory-mb	54272	27136

r6in.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r6in.12xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6in.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r6in.32xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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-Instances

r7a.x groß

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r7a.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7a.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

r7a.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7a.24xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r7a.48xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

R7g-Instances

r7g.xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r7g.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

r7g.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r7g.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

R7gd-Instances

r7gd.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7gd.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r7gd.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r7gd.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r7gd.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7i-Instances

r7i.xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7i.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r7i.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

r7i.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r7i.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r7i.24xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

R7iz-Instances

r7iz.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

r7iz.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7iz.4xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

r7iz.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7iz.12xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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

r7iz.32xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx6029m	-Xmx6029m
mapreduce.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.map.memory.mb	7536	7536
mapreduce.reduce.memory.mb	15072	15072
yarn.app.mapreduce.am.resource.mb	15072	15072

Konfigurationsoption	Standardwert	Mit HBase installiert
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource-memory-mb</code>	964608	30144

X1-Instances

x1.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
<code>mapreduce.map.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24116m</code>	<code>-Xmx24116m</code>
<code>mapreduce.map.memory.mb</code>	15072	15072
<code>mapreduce.reduce.memory.mb</code>	30144	30144
<code>yarn.app.mapreduce.am.resource.mb</code>	30144	30144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource-memory-mb</code>	964608	30144

x1.32xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx12109m	-Xmx12109m
mapreduce.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.map.memory.mb	15136	15136
mapreduce.reduce.memory.mb	30272	30272
yarn.app.mapreduce.am.resource.mb	30272	30272
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	30272
yarn.nodemanager.resource.memory-mb	1937408	30272

X1e-Instances

x1e.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx22682m	-Xmx22682m
mapreduce.java.opts	-Xmx45364m	-Xmx45364m
mapreduce.map.memory.mb	28352	28352
mapreduce.reduce.memory.mb	56704	56704
yarn.app.mapreduce.am.resource.mb	56704	56704

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	113408	0
yarn.nodemanager.resource.memory-mb	113408	0

x1e.2xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx23910m	-Xmx23910m
mapreduce.java.opts	-Xmx47820m	-Xmx47820m
mapreduce.map.memory.mb	29888	29888
mapreduce.reduce.memory.mb	59776	59776
yarn.app.mapreduce.am.resource.mb	59776	59776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	478208	0
yarn.nodemanager.resource.memory-mb	478208	0

x1e.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx24115m	-Xmx24115m
mapreduce.java.opts	-Xmx48230m	-Xmx48230m
mapreduce.map.memory.mb	30144	30144
mapreduce.reduce.memory.mb	60288	60288
yarn.app.mapreduce.am.resource.mb	60288	60288

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	0
yarn.nodemanager.resource.memory-mb	964608	0

x1e.16xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.java.opts	-Xmx48436m	-Xmx48436m
mapreduce.map.memory.mb	30272	30272
mapreduce.reduce.memory.mb	60544	60544
yarn.app.mapreduce.am.resource.mb	60544	60544
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	0
yarn.nodemanager.resource.memory-mb	1937408	0

x1e.32xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3883008	0
yarn.nodemanager.resource.memory-mb	3883008	0

X2gd-Instances

x2gd.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136
yarn.app.mapreduce.am.resource.mb	27136	27136

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

x2gd.2xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

x2gd.8xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.map.memory.mb	15360	15360
mapreduce.reduce.memory.mb	30720	30720
yarn.app.mapreduce.am.resource.mb	30720	30720

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

x2gd.12xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
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

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

x2idn-Instances

x2idn.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

x2idn.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx12442m	-Xmx12442m
mapreduce.java.opts	-Xmx24884m	-Xmx24884m
mapreduce.map.memory.mb	15552	15552
mapreduce.reduce.memory.mb	31104	31104
yarn.app.mapreduce.am.resource.mb	31104	31104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	31104
yarn.nodemanager.resource.memory-mb	1990656	31104

X2iedn-Instances

x2iedn.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx23347m	-Xmx23347m
mapreduce.java.opts	-Xmx46694m	-Xmx46694m
mapreduce.map.memory.mb	29184	29184
mapreduce.reduce.memory.mb	58368	58368
yarn.app.mapreduce.am.resource.mb	58368	58368

Konfigurationsoption	Standardwert	Mit installiert HBase
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	0
yarn.nodemanager.resource.memory-mb	116736	0

x2iedn.2xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	0
yarn.nodemanager.resource.memory-mb	491520	0

x2iedn.8xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx24781m	-Xmx24781m
mapreduce.java.opts	-Xmx49562m	-Xmx49562m
mapreduce.map.memory.mb	30976	30976
mapreduce.reduce.memory.mb	61952	61952
yarn.app.mapreduce.am.resource.mb	61952	61952

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	0
yarn.nodemanager.resource-memory-mb	991232	0

x2iedn.16xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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	1990656	0
yarn.nodemanager.resource-memory-mb	1990656	0

x2iedn.24xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
mapreduce.map.java.opts	-Xmx24918m	-Xmx24918m
mapreduce.java.opts	-Xmx49836m	-Xmx49836m
mapreduce.map.memory.mb	31147	31147
mapreduce.reduce.memory.mb	62294	62294
yarn.app.mapreduce.am.resource.mb	62294	62294
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	2990080	-32
yarn.nodemanager.resource.memory-mb	2990080	-32

x2iedn.32xlarge

Konfigurationsoption	Standardwert	Mit installiertem HBase
mapreduce.map.java.opts	-Xmx24934m	-Xmx24934m
mapreduce.java.opts	-Xmx49868m	-Xmx49868m
mapreduce.map.memory.mb	31168	31168
mapreduce.reduce.memory.mb	62336	62336
yarn.app.mapreduce.am.resource.mb	62336	62336

Konfigurationsoption	Standardwert	Mit installiertem HBase
<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

z1d-Instances

z1d.xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
<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

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

z1d.3xlarge

Konfigurationsoption	Standardwert	Mit HBase installiert
mapreduce.map.java.opts	-Xmx6007m	-Xmx6007m
mapreduce.java.opts	-Xmx12014m	-Xmx12014m
mapreduce.map.memory.mb	7509	7509
mapreduce.reduce.memory.mb	15018	15018
yarn.app.mapreduce.am.resource.mb	15018	15018

Konfigurationsoption	Standardwert	Mit HBase installiert
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource-memory-mb	90112	30040

z1d.6xlarge

Konfigurationsoption	Standardwert	Mit installiert HBase
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

Konfigurationsoption	Standardwert	Mit installiert HBase
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.resou rce.mb	16042	16042
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	385024	32100
yarn.nodemanager.resource.m emory-mb	385024	32100

Hadoop-Daemon-Konfigurationseinstellungen

Hadoop-Daemon-Einstellungen unterscheiden sich je nach dem EC2-Instance-Typ, den ein Cluster-Knoten verwendet. In der folgenden Tabelle sind die Standardkonfigurationseinstellungen für die einzelnen EC2 Instance-Typen in Clustern aufgeführt.

Zum Anpassen dieser Einstellungen verwenden Sie die `hadoop-env`-Konfigurationsklassifizierung. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Instance-Typen

- [C1-Instances](#)
- [C3-Instances](#)
- [C4-Instances](#)
- [C5-Instances](#)

- [C5a-Instances](#)
- [C5ad-Instances](#)
- [C5d-Instances](#)
- [C5n-Instances](#)
- [C6a-Instances](#)
- [C6g-Instances](#)
- [C6gd-Instances](#)
- [C6gn-Instances](#)
- [C6i-Instances](#)
- [C6id-Instances](#)
- [C6in-Instances](#)
- [c7a-Instances](#)
- [C7g-Instances](#)
- [C7gd-Instances](#)
- [C7gn-Instances](#)
- [C7i-Instances](#)
- [D2-Instances](#)
- [D3-Instances](#)
- [D3en-Instances](#)
- [G3-Instances](#)
- [G3s-Instances](#)
- [G4dn-Instances](#)
- [G5-Instances](#)
- [H1-Instances](#)
- [I2-Instances](#)
- [I3-Instances](#)
- [I3en-Instances](#)
- [I4g-Instances](#)

- [I4i-Instances](#)
- [I4gn-Instances](#)
- [I4gen-Instances](#)
- [M1-Instances](#)
- [M2-Instances](#)
- [M3-Instances](#)
- [M4-Instances](#)
- [M5-Instances](#)
- [M5a-Instances](#)
- [M5ad-Instances](#)
- [M5d-Instances](#)
- [M5dn-Instances](#)
- [M5n-Instances](#)
- [M5zn-Instances](#)
- [M6a-Instances](#)
- [M6g-Instances](#)
- [M6gd-Instances](#)
- [M6i-Instances](#)
- [M6id-Instances](#)
- [M6idn-Instances](#)
- [M6in-Instances](#)
- [M7a-Instances](#)
- [M7g-Instances](#)
- [M7gd-Instances](#)
- [M7i-Instances](#)
- [M7i-flex-Instances](#)
- [P2-Instances](#)
- [P3-Instances](#)

- [P5-Instances](#)
- [R3-Instances](#)
- [R4-Instances](#)
- [R5-Instances](#)
- [R5a-Instances](#)
- [R5ad-Instances](#)
- [r5b-Instances](#)
- [R5d-Instances](#)
- [R5dn-Instances](#)
- [R5n-Instances](#)
- [R6a-Instances](#)
- [R6g-Instances](#)
- [R6gd-Instances](#)
- [R6i-Instances](#)
- [R6id-Instances](#)
- [R6idn-Instances](#)
- [R6in-Instances](#)
- [R7a-Instances](#)
- [R7g-Instances](#)
- [R7gd-Instances](#)
- [r7i-Instances](#)
- [R7iz-Instances](#)
- [X1-Instances](#)
- [X1e-Instances](#)
- [X2gd-Instances](#)
- [x2idn-Instances](#)
- [X2iedn-Instances](#)
- [z1d-Instances](#)

C1-Instances

c1.medium

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	192
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	128
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

c1.xlarge

Parameter	Wert
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

C3-Instances

c3.xlarge

Parameter	Wert
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

c3.2xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C4-Instances

c4.large

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C5-Instances

c5.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C5a-Instances

c5a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C5ad-Instances

c5ad.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C5d-Instances

c5d.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C5n-Instances

c5n.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C6a-Instances

c6a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C6g-Instances

c6g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C6gd-Instances

c6gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C6gn-Instances

c6gn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

c6i.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C6id-Instances

c6id.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C6in-Instances

c6in.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

c7a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C7g-Instances

c7g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C7gd-Instances

c7gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

C7gn-Instances

c7gn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

c7i.x groß

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

D2-Instances

d2.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

D3-Instances

d3.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

D3en-Instances

d3en.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

G3-Instances

g3.4xgroß

Parameter	Wert
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

g3.8xgroß

Parameter	Wert
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

g3.16xgroß

Parameter	Wert
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

G3s-Instances

g3s.xgroß

Parameter	Wert
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

G4dn-Instances

g4dn.xgroß

Parameter	Wert
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.2xgroß

Parameter	Wert
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.4xgroß

Parameter	Wert
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.8xgroß

Parameter	Wert
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.12xgroß

Parameter	Wert
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.16xgroß

Parameter	Wert
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

G5-Instances

g5.xgroß

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

H1-Instances

h1.2xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

I2-Instances

i2.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i2.8xlarge

Parameter	Wert
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-Instances

i3.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

I3en-Instances

i3en.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

I4g-Instances

i4g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

i4i.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Im4gn-Instances

im4gn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Is4gen-Instances

is4gen.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M1-Instances

m1.small

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M2-Instances

m2.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M3-Instances

m3.xlarge

Parameter	Wert
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

Parameter	Wert
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

M4-Instances

m4.large

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

m5.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M5a-Instances

m5a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M5ad-Instances

m5ad.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M5d-Instances

m5d.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M5dn-Instances

m5dn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M5n-Instances

m5n.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M5zn-Instances

m5zn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

m6a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M6g-Instances

m6g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M6gd-Instances

m6gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

m6i.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M6id-Instances

m6id.x groß

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M6idn-Instances

m6idn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M6in-Instances

m6in.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

m6n.8xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

m7a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M7g-Instances

m7g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M7gd-Instances

m7gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

m7i.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

M7i-flex-Instances

m7i-flex.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

P2-Instances

p2.xgroß

Parameter	Wert
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

p2.8xgroß

Parameter	Wert
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

p2.16xgroß

Parameter	Wert
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

P3-Instances

p3.2xgroß

Parameter	Wert
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

p3.8xgroß

Parameter	Wert
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

p3.16xgroß

Parameter	Wert
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

P5-Instances

p5.48xlarge

Parameter	Wert
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

R3-Instances

r3.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

r4.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R5-Instances

r5.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R5a-Instances

r5a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R5ad-Instances

r5ad.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

r5b-Instances

r5b.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5b.2xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R5d-Instances

r5d.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R5dn-Instances

r5dn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R5n-Instances

r5n.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

r6a.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R6g-Instances

r6g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R6gd-Instances

r6gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

r6i.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R6id-Instances

r6id.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R6idn-Instances

r6idn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R6in-Instances

r6in.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

r7a.x groß

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R7g-Instances

r7g.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R7gd-Instances

r7gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

r7i.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

R7iz-Instances

r7iz.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

x1.16xlarge

Parameter	Wert
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

Parameter	Wert
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-Instances

x1e.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

X2gd-Instances

x2gd.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

x2idn.16xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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-Instances

x2iedn.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

z1d-Instances

z1d.xlarge

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

Parameter	Wert
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

HDFS Konfiguration

In der folgenden Tabelle werden die Standardparameter des Hadoop Distributed File System (HDFS) und ihre Einstellungen beschrieben. Sie können diese Werte mithilfe der `hdfs-site-`

Konfigurationsklassifizierung ändern. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Warning

1. Die Einstellung `dfs.replication` für Cluster mit weniger als vier Knoten kann zu HDFS Datenverlust führen, wenn ein einzelner Knoten ausfällt. Wenn Ihr Cluster über HDFS Speicher verfügt, empfehlen wir, den Cluster mit mindestens vier Kernknoten für Produktionsworkloads zu konfigurieren, um Datenverlust zu vermeiden.
2. Amazon EMR erlaubt Clustern nicht, Kernknoten nach unten zu skalieren `dfs.replication`. Bei `dfs.replication = 2` z. B. beträgt die Mindestanzahl von Core-Knoten 2.
3. Wenn Sie Verwaltete Skalierung oder Auto-Scaling verwenden oder die Größe Ihres Clusters manuell ändern möchten, empfehlen wir Ihnen, die Größe `dfs.replication` auf 2 oder höher einzustellen.

Parameter	Definition	Standardwert
<code>dfs.block.size</code>	Die Größe der HDFS Blöcke. Wenn mit Daten gearbeitet wird, die in gespeichert sind HDFS, entspricht die Teilungsgröße im Allgemeinen der Größe eines HDFS Blocks. Größere Größen bieten weniger Aufgabengranularität, belasten den Cluster NameNode jedoch auch weniger.	134217728 (128 MB)
<code>dfs.replication</code>	Die Anzahl der Kopien der einzelnen Blöcke, die dauerhaft gespeichert werden sollen. Amazon EMR legt diesen Wert auf der Grundlage der Anzahl der Kernknoten fest, mit denen der Cluster bereitgestellt wird. Passen Sie den Wert entsprechend Ihren Anforderungen an. Zum Überschreiben des Standardwerts verwenden Sie die <code>hdfs-site</code> -Klassifizierung.	1 für Cluster, die mit weniger als vier Core-Knoten bereitgestellt werden 2 für Cluster, die mit weniger als zehn Core-Knoten bereitgestellt werden

Parameter	Definition	Standardwert
		3 für alle anderen Cluster

Transparente Verschlüsselung HDFS bei Amazon EMR

Transparente Verschlüsselung wird mithilfe von HDFS Verschlüsselungszonen implementiert. Dabei handelt es sich um HDFS Pfade, die Sie definieren. Jede Verschlüsselungszone verfügt über ihren eigenen Schlüssel, der unter Verwendung der `hdfs-site`-Konfigurationsklassifikation auf dem Schlüssel-Server gespeichert wird.

Ab der EMR Amazon-Release-Version 4.8.0 können Sie EMR Amazon-Sicherheitskonfigurationen verwenden, um Datenverschlüsselungseinstellungen für Cluster einfacher zu konfigurieren. Sicherheitskonfigurationen bieten Einstellungen, um die Sicherheit für Daten während der Übertragung und Speicherung von Daten in Amazon Elastic Block Store (AmazonEBS) - Speichervolumen und EMRFS Daten in Amazon S3 zu aktivieren. Weitere Informationen finden Sie unter [Verschlüsseln von Daten bei der Übertragung und im Speicher](#) im Amazon EMR Management Guide.

Amazon EMR verwendet KMS standardmäßig Hadoop. Sie können jedoch auch ein anderes verwenden, KMS das den KeyProvider API Vorgang implementiert. Jede Datei in einer HDFS Verschlüsselungszone hat ihren eigenen eindeutigen Datenverschlüsselungsschlüssel, der mit dem Verschlüsselungszonenschlüssel verschlüsselt wird. HDFS Daten werden verschlüsselt end-to-end (im Speicher und während der Übertragung), wenn Daten in eine Verschlüsselungszone geschrieben werden, da Verschlüsselungs- und Entschlüsselungsaktivitäten nur auf dem Client stattfinden.

Sie können keine Dateien zwischen den Verschlüsselungszonen oder von einer Verschlüsselungszone nach unverschlüsselten Pfaden verschieben.

Der NameNode HDFS AND-Client interagiert während des Vorgangs mit Hadoop KMS (oder einer von KMS Ihnen konfigurierten Alternative). KeyProvider API Der KMS ist für das Speichern von Verschlüsselungsschlüsseln im Backing-Keystore verantwortlich. Amazon EMR bietet außerdem die Richtlinie „JCEUnlimited Strength“ an, sodass Sie Schlüssel in der gewünschten Länge erstellen können.

Weitere Informationen finden Sie unter [Transparente Verschlüsselung HDFS in](#) der Hadoop-Dokumentation.

Note

In Amazon EMR HTTPS ist KMS Over bei Hadoop KMS standardmäßig nicht aktiviert. Weitere Informationen zur Aktivierung von KMS Over HTTPS finden Sie in der [Hadoop-Dokumentation KMS](#).

Konfiguration HDFS transparenter Verschlüsselung

Sie können die transparente Verschlüsselung in Amazon konfigurieren, EMR indem Sie Schlüssel erstellen und Verschlüsselungszonen hinzufügen. Hierfür gibt es mehrere Möglichkeiten:

- Verwenden des EMR API Amazon-Konfigurationsvorgangs beim Erstellen eines Clusters
- Verwenden eines JAR Hadoop-Schritts mit `command-runner.jar`
- Anmelden am Master-Knoten des Hadoop-Clusters und verwenden der Befehlszeilen-Clients `hadoop key` und `hdfs crypto`
- Verwenden von REST APIs für KMS Hadoop und HDFS

Weitere Informationen zu finden Sie in der REST APIs entsprechenden Dokumentation für Hadoop KMS und. HDFS

Um Verschlüsselungszonen und ihre Schlüssel bei der Clustererstellung zu erstellen, verwenden Sie CLI

Die `hdfs-encryption-zones` Klassifizierung im API Konfigurationsvorgang ermöglicht es Ihnen, einen Schlüsselnamen und eine Verschlüsselungszone anzugeben, wenn Sie einen Cluster erstellen. Amazon EMR erstellt diesen Schlüssel in Hadoop KMS auf Ihrem Cluster und konfiguriert die Verschlüsselungszone.

- Erstellen Sie einen Cluster mit dem folgenden Befehl.

```
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

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

Manuelles Erstellen von Verschlüsselungszonen und deren Schlüssel auf dem Master-Knoten

1. Starten Sie Ihren Cluster mit einer EMR Amazon-Version größer als 4.1.0.
2. Stellen Sie mit dem Master-Knoten des Clusters eine ConnectSSH.
3. Erstellen Sie einen Schlüssel in HadoopKMS.

```
$ hadoop key create path2_key
path2_key has been successfully created with options Options{cipher='AES/CTR/
NoPadding', bitLength=256, description='null', attributes=null}.
KMSClientProvider[http://ip-x-x-x-x.ec2.internal:16000/kms/v1/] has been updated.
```

Important

Hadoop KMS verlangt, dass Ihre Schlüsselnamen in Kleinbuchstaben geschrieben werden. Wenn Sie einen Schlüssel mit Großbuchstaben verwenden, wird der Start Ihres Clusters fehlschlagen.

4. Erstellen Sie den Pfad zur Verschlüsselungszone in HDFS

```
$ hadoop fs -mkdir /myHDFSPath2
```

5. Machen Sie den HDFS Pfad mithilfe des von Ihnen erstellten Schlüssels zu einer Verschlüsselungszone.

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2
Added encryption zone /myHDFSPath2
```

Um Verschlüsselungszonen und ihre Schlüssel manuell zu erstellen, verwenden Sie AWS CLI

- Fügen Sie Schritte hinzu, um die KMS Schlüssel und Verschlüsselungszonen manuell mit dem folgenden Befehl zu erstellen.

```
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

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Überlegungen zur HDFS transparenten Verschlüsselung

Eine bewährte Methode ist das Erstellen einer Verschlüsselungszone für jede Anwendung, in der möglicherweise Dateien geschrieben werden. Sie können auch alles verschlüsseln, HDFS indem Sie die `hdfs-encryption-zones` Klassifizierung in der Konfiguration verwenden API und den Stammpfad (`/`) als Verschlüsselungszone angeben.

Hadoop-Schlüsselverwaltungsserver (KMS)

[Hadoop KMS](#) ist ein Schlüsselverwaltungsserver, der die Möglichkeit bietet, kryptografische Dienste für Hadoop-Cluster zu implementieren, und der als Schlüsselanbieter für dienen kann. [Transparente Verschlüsselung HDFS bei Amazon EMR](#) Hadoop KMS in Amazon EMR ist standardmäßig installiert und aktiviert, wenn Sie beim Starten eines EMR Clusters die Hadoop-Anwendung auswählen. Hadoop speichert die Schlüssel selbst KMS nicht, es sei denn, es handelt sich um temporäres Caching. Hadoop fungiert als Proxy zwischen dem Schlüsselanbieter und dem Client-Trustee zu einem Backing-Keystore — es ist kein Keystore. Der Standard-Keystore, der für Hadoop erstellt wird, ist die Java Cryptography Extension (`JCEKS`). Die Richtlinie „JCEUnlimited Strength“ ist ebenfalls enthalten, sodass Sie Schlüssel mit der gewünschten Länge erstellen können. Hadoop unterstützt KMS auch eine Reihe ACLs dieser Funktionen zur Steuerung des Zugriffs auf Schlüssel und Schlüsseloperationen unabhängig von anderen Client-Anwendungen, wie HDFS z. Die Standardschlüssellänge in Amazon EMR ist 256 Bit.

Verwenden Sie zur Konfiguration von Hadoop die `hadoop-kms-site` Klassifizierung `KMS`, um die Einstellungen zu ändern. Zur Konfiguration ACLs verwenden Sie die Klassifizierung `kms-acls`.

[Weitere Informationen finden Sie in der Hadoop-Dokumentation. KMS](#) Hadoop KMS wird in der transparenten Hadoop-Verschlüsselung HDFS verwendet. Weitere Informationen zur HDFS transparenten Verschlüsselung finden Sie im Thema [HDFS transparente Verschlüsselung](#) in der Apache Hadoop-Dokumentation.

Note

In Amazon EMR HTTPS ist KMS Over bei Hadoop KMS standardmäßig nicht aktiviert. Informationen zur Aktivierung von KMS Over HTTPS finden Sie in der [Hadoop-Dokumentation KMS](#).

⚠ Important

Hadoop KMS verlangt, dass Ihre Schlüsselnamen in Kleinbuchstaben geschrieben werden. Wenn Sie einen Schlüssel mit Großbuchstaben verwenden, wird der Start Ihres Clusters fehlschlagen.

Konfiguration von Hadoop KMS in Amazon EMR

Wenn Sie die EMR Amazon-Release-Version 4.6.0 oder höher verwenden, `kms-http-port` ist das 9700 und `kms-admin-port` ist 9701.

Sie können Hadoop KMS bei der Clustererstellung mithilfe der Konfiguration API für EMR Amazon-Releases konfigurieren. Im Folgenden sind die für Hadoop verfügbaren Klassifizierungen von Konfigurationsobjekten aufgeführt: KMS

Klassifizierungen der Hadoop-Konfiguration KMS

Klassifizierung	Dateiname
hadoop-kms-site	kms-site.xml
hadoop-kms-acls	kms-acls.xml
hadoop-kms-env	kms-env.sh
hadoop-kms-log4j	kms-log4j.properties

So richten Sie Hadoop KMS ACLs mit dem ein CLI

- Erstellen Sie mit dem folgenden Befehl einen Cluster KMS ACLs mit Hadoop:

```
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

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-acls",
    "Properties": {
      "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
      "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
    }
  }
]
```

Um den KMS Hadoop-Cache zu deaktivieren, verwenden Sie den CLI

- Erstellen Sie mit dem folgenden Befehl einen Cluster `false`, für den Hadoop auf KMS `hadoop.kms.cache.enable` eingestellt ist:

```
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

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.cache.enable": "false"
    }
  }
]
```

Um Umgebungsvariablen im **kms-env.sh** Skript festzulegen, verwenden Sie den CLI

- Ändern der Einstellungen in **kms-env.sh** über die **hadoop-kms-env**-Konfiguration. Erstellen Sie mit dem folgenden Befehl einen Cluster KMS mit Hadoop:

```
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

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

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": [
    ]
}
]
}
]

```

Informationen zur Konfiguration von Hadoop KMS finden Sie in der [KMS Hadoop-Dokumentation](#).

HDFS transparente Verschlüsselung auf EMR Clustern mit mehreren Master-Knoten

[Apache Ranger](#) KMS wird in einem EMR Amazon-Cluster mit mehreren Primärknoten für eine transparente Verschlüsselung verwendet. HDFS

Apache Ranger KMS speichert seinen Root-Schlüssel und seine Encryption Zone (EZ) -Schlüssel in Ihrem Amazon RDS für einen EMR Amazon-Cluster mit mehreren Primärknoten. Um die transparente Verschlüsselung in HDFS einem EMR Amazon-Cluster mit mehreren Primärknoten zu aktivieren, müssen Sie die folgenden Konfigurationen angeben.

- Amazon RDS oder Ihre eigene My SQL Server-Verbindung URL zum Speichern des KMS Ranger-Root-Schlüssels und des EZ-Schlüssels
- Benutzername und Passwort für My SQL
- Passwort für den KMS Ranger-Root-Schlüssel
- PEMZertifizierungsstellendatei (CA) für die SSL Verbindung zu My Server SQL

Sie können diese Konfigurationen mithilfe der `ranger-kms-dbks-site`-Klassifizierung und der `ranger-kms-db-ca`-Klassifizierung bereitstellen, wie im folgenden Beispiel veranschaulicht.

```

[
  {
    "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"
  }
}
]

```

Im Folgenden finden Sie Klassifizierungen von Konfigurationsobjekten für Apache KMS Ranger.

Klassifizierungen der Hadoop-Konfiguration KMS

Klassifizierung	Beschreibung
ranger-kms-dbks-site	Ändern Sie die Werte in der Datei dbks-site.xml von Ranger. KMS
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger. KMS
ranger-kms-env	Ändern Sie Werte in der KMS Ranger-Umgebung.
ranger-kms-log4j	Ändern Sie die Werte in der Datei kms-log4j.properties von Ranger. KMS
ranger-kms-db-ca	Ändern Sie die Werte für die CA-Datei auf S3 für Meine Verbindung mit Ranger. SQL SSL KMS

Überlegungen

- Es wird dringend empfohlen, Ihre RDS Amazon-Instance zu verschlüsseln, um die Sicherheit zu verbessern. Weitere Informationen finden Sie unter [Überblick über die Verschlüsselung von RDS Amazon-Ressourcen](#).
- Es wird dringend empfohlen, für jeden EMR Amazon-Cluster mit mehreren Primärknoten eine separate My SQL Database zu verwenden, um eine hohe Sicherheitsleistung zu gewährleisten.

- Um die transparente Verschlüsselung in HDFS einem EMR Amazon-Cluster mit mehreren Primärknoten zu konfigurieren, müssen Sie die `hdfs-encryption-zones` Klassifizierung bei der Erstellung des Clusters angeben. Andernfalls KMS wird Ranger nicht konfiguriert oder gestartet. Die Neukonfiguration der `hdfs-encryption-zones` Klassifizierung oder einer der KMS Hadoop-Konfigurationsklassifizierungen auf einem laufenden Cluster wird auf einem EMR Amazon-Cluster mit mehreren Primärknoten nicht unterstützt.

Erstellen oder Ausführen einer Hadoop-Anwendung

Themen

- [Erstellen von Binärdateien mit Amazon EMR](#)
- [Daten per Streaming verarbeiten](#)
- [Verarbeiten Sie Daten mit einem benutzerdefinierten JAR](#)

Erstellen von Binärdateien mit Amazon EMR

Sie können Amazon EMR als Build-Umgebung zum Kompilieren von Programmen für die Verwendung in Ihrem Cluster nutzen. Programme, die Sie mit Amazon EMR verwenden, müssen auf einem System kompiliert werden, das dieselbe Linux-Version wie Amazon EMR verwendet. Bei einer 32 Bit-Version sollten Sie die Kompilierung auf einem 32 Bit-Computer oder mit aktivierter 32 Bit-Cross-Compilation-Option durchgeführt haben. Bei einer 64 Bit-Version müssen Sie die Kompilierung auf einem 64 Bit-Computer oder mit aktivierter 64 Bit-Cross-Compilation-Option durchgeführt haben. Weitere Informationen zu EC2-Instance-Versionen finden Sie unter [Planen und Konfigurieren von EC2-Instances](#) im Verwaltungshandbuch für Amazon EMR. Unterstützte Programmiersprachen sind unter anderem C++, Python und C #.

In der folgenden Tabelle werden die Schritte für das Erstellen und Testen Ihrer Anwendung mit Amazon EMR aufgeführt.

Prozess zum Erstellen eines Moduls

- 1 Stellen Sie eine Verbindung mit dem Master-Knoten Ihres Clusters her.
- 2 Kopieren Sie Quelldateien auf den Master-Knoten.
- 3 Erstellen Sie Binärdateien mit den erforderlichen Optimierungen.

4 Kopieren Sie die Binärdateien vom Hauptknoten zu Amazon S3.

Die Einzelheiten für jeden dieser Schritte finden Sie in den folgenden Abschnitten.

So verbinden Sie sich mit dem Master-Knoten des Clusters

- Befolgen Sie die Anweisungen unter [Verbindung mit dem Hauptknoten mit SSH](#) im Verwaltungshandbuch für Amazon EMR.

So kopieren Sie Quelldateien auf den Master-Knoten

1. Platzieren Sie Ihre Quelldateien in einem Amazon-S3-Bucket. Weitere Informationen zum Erstellen von Buckets und zum Verschieben von Daten nach Amazon S3 finden Sie im [Benutzerhandbuch für den Amazon Simple Storage Service](#).
2. Erstellen Sie einen Ordner für Ihre Quelldateien auf Ihrem Hadoop-Cluster, indem Sie einen Befehl wie den folgenden eingeben:

```
mkdir SourceFiles
```

3. Kopieren Sie Ihre Quelldateien von Amazon S3, auf den Hauptknoten, indem Sie einen Befehl wie den folgenden eingeben:

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

Erstellen von Binärdateien mit den erforderlichen Optimierungen

Die Erstellung Ihrer Binärdateien hängt von vielen Faktoren ab. Befolgen Sie beim Einrichten und Konfigurieren Ihrer Umgebung die Anweisungen zu Ihren Build-Tools. Sie können Hadoop-System-Specification-Befehle nutzen, um Cluster-Informationen zur Definition der Installation Ihrer Build-Umgebung abzurufen.

So ermitteln Sie Systemspezifikationen

- Verwenden Sie die folgenden Befehle, um die zum Erstellen Ihrer Binärdateien verwendete Architektur abzurufen.
 - a. Um die Debian-Version abzurufen, geben Sie den folgenden Befehl ein:

```
master$ cat /etc/issue
```

Die Ausgabe sieht folgendermaßen oder ähnlich aus.

```
Debian GNU/Linux 5.0
```

- b. Um den öffentlichen DNS-Namen und die Prozessorgröße abzurufen, geben Sie den folgenden Befehl ein:

```
master$ uname -a
```

Die Ausgabe sieht folgendermaßen oder ähnlich aus.

```
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. Um die Prozessorgeschwindigkeit abzurufen, geben Sie den folgenden Befehl ein:

```
master$ cat /proc/cpuinfo
```

Die Ausgabe sieht folgendermaßen oder ähnlich aus.

```
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  
...
```

Sobald Ihre Binärdateien erstellt sind, können Sie die Dateien in Amazon S3 kopieren.

So kopieren Sie die Binärdateien vom Hauptknoten zu Amazon S3

- Geben Sie den folgenden Befehl ein, um die Binärdateien in Ihren Amazon-S3-Bucket zu kopieren:

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

Daten per Streaming verarbeiten

Hadoop-Streaming ist ein im Lieferumfang von Hadoop enthaltenes Hilfsprogramm, mit dem Sie MapReduce ausführbare Dateien in anderen Sprachen als Java entwickeln können. Streaming wird in Form einer JAR Datei implementiert, sodass Sie es wie eine JAR Standarddatei über Amazon EMR API oder die Befehlszeile ausführen können.

In diesem Abschnitt wird beschrieben, wie Sie Streaming mit Amazon verwendenEMR.

Note

Apache Hadoop Streaming ist ein unabhängiges Tool. Daher werden seine Funktionen und Parameter hier nicht beschrieben. Weitere Informationen zum Hadoop-Streaming finden Sie [unter http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html](http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html).

Verwenden des Hadoop-Streaming-Dienstprogramms

In diesem Abschnitt wird die Verwendung des Hadoop-Streaming-Dienstprogramms beschrieben.

Hadoop-Prozess

- 1 Schreiben Sie Ihre ausführbare Mapper- und Reducer-Datei in der Programmiersprache Ihrer Wahl.

Folgen Sie den Anweisungen in der Dokumentation zu Hadoop, um die ausführbaren Dateien zu schreiben. Die Programme sollten ihre Eingaben über die Standardeingabe lesen und die Daten über die Standardausgabedaten ausgeben. Standardmäßig stellt jede Zeile der Eingabe/Ausgabe einen Datensatz dar. Der erste Tabulator jeder Zeile dient als Trennzeichen zwischen Schlüssel und Wert.
- 2 Testen Sie Ihre ausführbaren Dateien lokal und laden Sie sie in Amazon S3 hoch.
- 3 Verwenden Sie die EMR Amazon-Befehlszeilenschnittstelle oder die EMR Amazon-Konsole, um Ihre Anwendung auszuführen.

Jedes Mapper-Skript startet als separater Prozess im Cluster. Jede ausführbare Reducer-Datei wandelt die Ausgabe der ausführbaren Mapper-Datei in die Datenausgabe des Auftragverlaufs um.

Die Parameter `input`, `output`, `mapper` und `reducer` sind für die meisten Streaming-Anwendungen erforderlich. Die folgende Tabelle beschreibt diese und andere (optionale) Parameter.

Parameter	Beschreibung	Erforderlich
<code>-input</code>	<p>Speicherort der Eingabedaten in Amazon S3.</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Einschränkung:URI. Wenn kein Protokoll angegeben wurde, wird das Standarddateisystem des Clusters verwendet.</p>	Ja
<code>-output</code>	<p>Ort auf Amazon S3, an den Amazon die verarbeiteten Daten EMR hochlädt.</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Einschränkung: URI</p> <p>Standard: Wenn kein Standort angegeben ist, EMR lädt Amazon die Daten an den von <code>input</code> angegebenen Standort hoch.</p>	Ja
<code>-mapper</code>	<p>Name der ausführbaren Mapper-Datei.</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p>	Ja
<code>-reducer</code>	<p>Name der ausführbaren Reducer-Datei.</p> <p>Typ: Zeichenfolge</p>	Ja

Parameter	Beschreibung	Erforderlich
	Standard: keiner	
-cacheFile	<p>Ein Amazon-S3-Speicherort mit Dateien, die Hadoop in Ihr lokales Arbeitsverzeichnis kopieren soll (in erster Linie zur Verbesserung der Leistung).</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Einschränkungen: [URI] # [Name des Symlinks, der im Arbeitsverzeichnis erstellt werden soll]</p>	Nein
-cacheArchive	<p>JARDatei zum Extrahieren in das Arbeitsverzeichnis</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Einschränkungen: [URI] # [Name des Symlink-Verzeichnisses, das im Arbeitsverzeichnis erstellt werden soll]</p>	Nein
-combiner	<p>Kombiniert Ergebnisse</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Einschränkungen: Java-Klassenname</p>	Nein

Das folgende Codebeispiel ist eine ausführbare Mapper-Datei, die in Python geschrieben ist. Dieses Skript ist Teil der WordCount Beispielanwendung.

```
#!/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)
```

Einen Streaming-Schritt übermitteln

In diesem Abschnitt werden die Grundlagen der Übermittlung eines Streaming-Schritts an einen Cluster behandelt. Eine Streaming-Anwendung liest Eingaben aus der Standardeingabe und führt dann ein Skript oder eine ausführbare Datei (der Mapper) für den Eingang aus. Das Ergebnis der einzelnen Eingaben wird lokal gespeichert, normalerweise auf einer Hadoop Distributed File System (HDFS) -Partition. Nachdem alle Eingaben vom Mapper verarbeitet wurden, verarbeitet ein zweites Skript oder eine ausführbare Datei (der Reducer) die Ergebnisse des Mappers. Die Ergebnisse aus dem Reducer werden an die Standardausgabe gesendet. Sie können eine Reihe von Streaming-Schritten verketteten, wobei die Ausgaben eines Schritts zur Eingabe eines weiteren Schritts werden.

Der Mapper und der Reducer können als Datei oder als Java-Klasse referenziert werden. Sie können den Mapper und den Reducer in jeder der unterstützten Sprachen implementieren, einschließlich Ruby, Perl/PHP, Python oder Bash.

Einen Streaming-Schritt mithilfe der Konsole übermitteln

In diesem Beispiel wird beschrieben, wie Sie mit der EMR Amazon-Konsole einen Streaming-Schritt an einen laufenden Cluster senden.

So übermitteln Sie einen Streaming-Schritt

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie unter Cluster List (Clusterliste) den Namen des Clusters aus.
3. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
4. Gehen Sie im Dialogfeld Add Step folgendermaßen vor:

- Wählen Sie unter Step type (Schritttyp) die Option Streaming program (Streaming-Programm) aus.
 - Übernehmen Sie unter Name den Standardnamen (Streaming program) oder geben Sie einen neuen Namen ein.
 - Geben Sie für Mapper den Speicherort Ihrer Mapper-Klasse in Hadoop oder einen S3-Bucket ein, in dem sich die ausführbare Mapper-Datei (z. B. ein Python-Programm) befindet. Der Pfadwert muss das folgende Format haben *BucketName/path/MapperExecutable*.
 - Geben Sie für Reducer den Speicherort Ihrer Reducer-Klasse in Hadoop oder einen S3-Bucket ein, in dem sich die ausführbare Reducer-Datei (z. B. ein Python-Programm) befindet. Der Pfadwert muss das folgende Format haben *BucketName/path/ReducerExecutable*. Amazon EMR unterstützt das spezielle Aggregate-Schlüsselwort. Weitere Informationen finden Sie in der Aggregate-Bibliothek von Hadoop.
 - Geben Sie unter Input S3 location (S3-Eingabespeicherort) den Speicherort Ihrer Eingabedaten ein oder browsen zum Speicherort der Daten.
 - Unter S3-Ausgabespeicherort geben Sie den Namen Ihres Amazon-S3-Ausgabe-Buckets ein oder navigieren zu dem entsprechenden S3-Bucket.
 - Lassen Sie unter Arguments (Parameter) das Feld leer.
 - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
5. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
 6. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie das Symbol Refresh über der Spalte "Actions" aus, um den Status zu aktualisieren.

AWS CLI

Diese Beispiele zeigen, wie Sie mit AWS CLI dem einen Cluster erstellen und einen Streaming-Schritt einreichen können.

Um einen Cluster zu erstellen und einen Streaming-Schritt einzureichen, verwenden Sie den AWS CLI

- Um einen Cluster zu erstellen und einen Streaming-Schritt mit dem zu senden AWS CLI, geben Sie den folgenden Befehl ein und ersetzen Sie *myKey* mit dem Namen Ihres EC2 key pair. Beachten Sie, dass Ihr Argument für `--files` der Amazon-S3-Pfad zum Speicherort Ihres

Skripts und die Argumente für `-mapper` und `-reducer` die Namen der jeweiligen Skriptdateien sein sollten.

```
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

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

Wenn Sie die Instance-Anzahl ohne den Parameter `--instance-groups` angeben, wird ein einzelner Master-Knoten gestartet. Die verbleibenden Instances werden als Core-Knoten aufgerufen. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

Note

Wenn Sie noch nicht die standardmäßige EMR Amazon-Servicerolle und das EC2 Instanzprofil erstellt haben, geben Sie `aws emr create-default-roles` um sie zu erstellen, bevor Sie den `create-cluster` Unterbefehl eingeben.

Weitere Informationen zur Verwendung von EMR Amazon-Befehlen in der AWS CLI finden Sie unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Verarbeiten Sie Daten mit einem benutzerdefinierten JAR

Ein Custom JAR führt ein kompiliertes Java-Programm aus, das Sie auf Amazon S3 hochladen können. Sie sollten das Programm für die Version von Hadoop kompilieren, die Sie starten möchten, und einen `CUSTOM_JAR` Schritt an Ihren EMR Amazon-Cluster senden. Weitere Informationen zum Kompilieren einer JAR Datei finden Sie unter [Erstellen von Binärdateien mit Amazon EMR](#).

Weitere Informationen zum Erstellen einer MapReduce Hadoop-Anwendung finden Sie im [MapReduce Tutorial](#) in der Apache Hadoop-Dokumentation.

Themen

- [Reichen Sie einen benutzerdefinierten Schritt ein JAR](#)

Reichen Sie einen benutzerdefinierten Schritt ein JAR

Ein Custom JAR führt ein kompiliertes Java-Programm aus, das Sie auf Amazon S3 hochladen können. Sie sollten das Programm für die Version von Hadoop kompilieren, die Sie starten möchten, und einen CUSTOM_JAR Schritt an Ihren EMR Amazon-Cluster senden. Weitere Informationen zum Kompilieren einer JAR Datei finden Sie unter [Erstellen von Binärdateien mit Amazon EMR](#).

Weitere Informationen zum Erstellen einer MapReduce Hadoop-Anwendung finden Sie im [MapReduce Tutorial](#) in der Apache Hadoop-Dokumentation.

In diesem Abschnitt werden die Grundlagen für das Einreichen eines benutzerdefinierten JAR Schritts bei Amazon behandelt. Wenn Sie einen benutzerdefinierten JAR Schritt einreichen, können Sie ein Skript schreiben, um Ihre Daten mit der Programmiersprache Java zu verarbeiten.

Reichen Sie einen benutzerdefinierten JAR Schritt mit der Konsole ein

In diesem Beispiel wird beschrieben, wie Sie mit der EMR Amazon-Konsole einen benutzerdefinierten JAR Schritt an einen laufenden Cluster senden.

Um einen benutzerdefinierten JAR Schritt mit der Konsole einzureichen

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie unter Cluster List (Clusterliste) den Namen des Clusters aus.
3. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
4. Gehen Sie im Dialogfeld Schritt hinzufügen folgendermaßen vor:
 - Wählen Sie als Schritttyp die Option Benutzerdefiniert aus JAR.
 - Akzeptieren Sie für Name den Standardnamen (BenutzerdefiniertJAR), oder geben Sie einen neuen Namen ein.

- Geben Sie für den JARS3-Speicherort den Speicherort Ihrer JAR Datei ein, oder suchen Sie danach. JARSpeicherort, vielleicht ein Pfad zu S3 oder eine vollständig qualifizierte Java-Klasse im Klassenpfad..
 - Geben Sie unter Arguments (Argumente) Ihre Argumente als durch Leerzeichen getrennte Zeichenfolgen ein oder lassen Sie das Feld leer.
 - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
5. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
 6. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie das Symbol Refresh über der Spalte "Actions" aus, um den Status zu aktualisieren.

Einen Cluster starten und einen benutzerdefinierten JAR Schritt einreichen mit AWS CLI

Um einen Cluster zu starten und einen benutzerdefinierten JAR Schritt einzureichen mit dem AWS CLI

Um einen Cluster zu starten und einen benutzerdefinierten JAR Schritt mit dem zu senden AWS CLI, geben Sie den `create-cluster` Unterbefehl mit dem `--steps` Parameter ein.

- Um einen Cluster zu starten und einen benutzerdefinierten JAR Schritt einzureichen, geben Sie den folgenden Befehl ein: replace *myKey* mit dem Namen Ihres EC2 key pair und ersetzen *mybucket* mit Ihrem Bucket-Namen.

```
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","pathtooutputbuck
```

Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Wenn Sie die Instance-Anzahl ohne den `--instance-groups`-Parameter angeben, wird ein einzelner Primärknoten gestartet. Die verbleibenden Instances werden dabei als Core-Knoten gestartet. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

Note

Wenn Sie noch nicht die standardmäßige EMR Amazon-Servicerolle und das EC2 Instanzprofil erstellt haben, geben Sie ein, `aws emr create-default-roles` um sie zu erstellen, bevor Sie den `create-cluster` Unterbefehl eingeben.

Weitere Informationen zur Verwendung von EMR Amazon-Befehlen in der AWS CLI finden Sie unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Abhängigkeiten von Drittanbietern

Manchmal kann es notwendig sein, es in den MapReduce Klassenpfad aufzunehmen, damit es mit Ihrem Programm JARs verwendet werden kann. Sie haben dafür zwei Möglichkeiten:

- Fügen Sie `--libjars s3://URI_to_JAR` in den Schrittoptionen für die Vorgehensweise in [Einen Cluster starten und einen benutzerdefinierten JAR Schritt einreichen mit AWS CLI](#) ein.
- Starten Sie den Cluster mit einer geänderten `mapreduce.application.classpath`-Einstellung in `mapred-site.xml`. Verwenden Sie die `mapred-site`-Konfigurationsklassifizierung. Um den Cluster mit dem Schritt `using` zu erstellen AWS CLI, würde das wie folgt aussehen:

```
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"
}
]
```

Die durch Kommas getrennte Pfadliste sollte für jede Aufgabe an den JVM Klassenpfad angehängt werden.

Wiederhergestellte Objekte lesen

Mit EMR Amazon-Version 7.2.0 und höher können Sie wiederhergestellte Glacier-Objekte vom S3-Speicherort der Tabelle mit dem S3A Protokoll lesen. Engines in früheren Versionen unterscheiden nicht zwischen Glacier- und Glacier Deep Archive-Dateien, was bedeutet, dass Sie eine erhalten würden, `AmazonS3Exception` wenn Sie versuchen würden, mit auf eine in Bearbeitung befindliche Glacier-Datei zuzugreifen. S3A Bei diesem Lesevorgang werden archivierte Glacier-Dateien ignoriert, wenn sie noch wiederhergestellt werden. Verwenden Sie die Einstellung `fs.s3a.glacier.read.restored.objects`, um dieses Verhalten zu aktivieren. Diese Einstellung kann drei Werte haben:

- `READ_ALL` — Dieser Wert gibt an, dass Amazon die von Amazon S3 abgerufenen Speicherklassen nicht berücksichtigen EMR sollte. Dies ist das Standardverhalten.
- `SKIP_ALL_GLACIER` — Dieser Wert gibt an, dass Amazon alle S3-Objekte ignorieren EMR sollte, die mit der Glacier-Speicherklasse gekennzeichnet sind, und alle anderen Objekte abrufen sollte. Dies ist das Standardverhalten für Amazon Athena in Bezug auf Glacier-Objekte.
- `READ_RESTORED_GLACIER_OBJECTS` — Dieser Wert gibt an, dass Amazon den wiederhergestellten Status des Glacier-Objekts überprüfen EMR soll. Wenn Amazon das Objekt wiederherstellen EMR kann, können Sie es wie ein normales S3-Objekt lesen. Andernfalls EMR ignoriert Amazon das Objekt von S3 Glacier

Beispiele

Spark

Verwenden Sie die folgende Konfiguration, um wiederhergestellte Objekte zu lesen, wenn Sie Spark verwenden:

```
--conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Wenn Sie Spark-Sql verwenden, verwenden Sie stattdessen die folgende Konfiguration:

```
spark-sql --conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Flink

Wenn Sie Flink verwenden, können Sie die Konfiguration in der Datei festlegen `flink-conf.yaml`

```
fs.s3a.glacier.read.restored.objects: <value>
```

Sie können auch die `flink-conf` Klassifizierung festlegen:

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "fs.s3a.glacier.read.restored.objects": "<value>"
    }
  }
]
```

Hive

Wenn Sie Hive verwenden, legen Sie die Konfiguration in der `hive-site.xml` Datei fest.

```
<property>
  <name>fs.s3a.glacier.read.restored.objects</name>
  <value><value></value>
</property>
```

Sie können den Hive auch verwendenCLI, um die Eigenschaft festzulegen: `--hiveconf`

```
hive --hiveconf fs.s3a.glacier.read.restored.objects=<value>
```

Überlegungen

Beachten Sie beim Lesen wiederhergestellter Objekte aus S3 Glacier die folgenden Überlegungen:

- Sie können wiederhergestellte Objekte nur lesen, wenn Sie das S3A Schema oder die verwenden `S3AFileSystem`, um auf die Daten zuzugreifen.
- Wenn Sie ein wiederhergestelltes Glacier-Objekt lesen, stellt Amazon das Objekt selbst EMR nicht wieder her. Dazu müssen Sie das AWS CLI oder das verwenden AWS SDK.

Aktivieren Sie die Erkennung ungleichmäßiger Speicherzugriffe für Container YARN

Mit EMR Amazon-Versionen 6.x und höher können Sie Non-Uniform Memory Access (NUMA) für die Mehrfachverarbeitung Ihrer Daten auf Clustern verwenden. NUMA ist ein Entwurfsmuster für Computerspeicher, bei dem der Prozessor schneller auf seinen eigenen lokalen Speicher zugreifen kann als auf Speicher auf einem anderen Prozessor oder auf den Speicher, der von mehreren Prozessoren gemeinsam genutzt wird. YARN Container weisen eine bessere Leistung auf NUMA, da sie sich an einen bestimmten NUMA Knoten binden können, der alle nachfolgenden Speicherzuweisungen bedient. Dadurch wird die Häufigkeit reduziert, mit der Ihr Cluster auf den Remotespeicher zugreifen muss.

Sie können die NUMA Unterstützung für YARN Container aktivieren, wenn es sich bei dem Worker-Node-Computer um einen Computer mit mehreren NUMA Knoten handelt. Führen Sie den folgenden Befehl aus, um zu überprüfen, ob es sich bei einem NUMA-Worker-Knoten um einen Single-NUMA- oder Multi-Knoten handelt.

```
lscpu | grep -i numa  
NUMA node(s): 2
```

Im Allgemeinen haben Instances, die größer als 12x sind, zwei NUMA-Knoten. Dies gilt nicht für Metal-Instances..

Um die Erkennung von NUMA YARN Containern zu aktivieren

1. Verwenden Sie die folgende `yarn-site` Konfiguration in Ihrem Amazon EMR 6.x-Cluster.

```
[  
  {  
    "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. Stellen Sie die folgende Bootstrap-Aktion in Ihrem Cluster bereit.

```

#!/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. Jeder Container muss NUMA kennen. Sie können die virtuelle Java-Maschine (JVM) in jedem Container mit einem NUMA Flag benachrichtigen. Um beispielsweise die zur Verwendung NUMA in einem MapReduce Job JVM zu benachrichtigen, fügen Sie die folgenden Eigenschaften hinzu `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>
```

- Um zu überprüfen, ob NUMA aktiviert sind, durchsuchen Sie eine der NodeManager-Protokolldateien mit dem folgenden Befehl.

```
grep "NUMA resources allocation is enabled," *
```

Um zu überprüfen, NodeManager ob einem Container NUMA Knotenressourcen zugewiesen wurden, durchsuchen NodeManager Sie das Protokoll mit dem folgenden Befehl und ersetzen `<container_id>` Sie es durch Ihre eigene Container-ID.

```
grep "NUMA node" | grep <container_id>
```

Hadoop-Versionsverlauf

In der folgenden Tabelle sind die in jeder Release-Version von Amazon EMR enthaltenen Hadoop-Versionen sowie die mit der Anwendung installierten Komponenten aufgeführt. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Hadoop-Versionsinformationen

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
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-resour

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
		cemanager, hadoop-yarn-timeline-server
emr-5.36,2	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.1.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-7.0.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.15.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.14.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.13.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.12.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.11.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.11.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.10.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.10.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.9.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.8.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.8.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.7.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.36.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.6.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.35.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.5.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.4.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.3.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.3.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.2.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.2.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.1.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.1.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.0.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.0.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.34.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.33.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.33.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.32.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.32.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.1	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.31.0	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.30.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
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
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

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.28.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.27.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.27.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.26.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.25.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.24.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.24.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.23.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.23.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.22.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.21.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.21.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.20.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.20.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.19.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.19.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.18.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.18.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.2	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.17.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.16.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.16.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.15.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.15.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.14.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.14.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.13.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.3	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.12.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.12.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.11.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.11.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.10.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.10.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.9.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.8.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.7.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.7.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.6.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.6.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.5.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.5.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.5.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.3.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.2.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.2.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.0.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.0.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.6	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.9.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
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-resource-manager
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-resource-manager
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-resource-manager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.7.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.6.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.5.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.4.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.3.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.2.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR Release-Etikett	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.1.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.0.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Hadoop-Versionshinweise nach Version

Themen

- [Amazon EMR 7.2.0 — Hadoop-Versionshinweise](#)
- [Amazon EMR 6.6.0 — Hadoop-Versionshinweise](#)

Amazon EMR 7.2.0 — Hadoop-Versionshinweise

Amazon EMR 7.2.0 — Hadoop-Änderungen

Typ	Beschreibung
-----	--------------

Typ	Beschreibung
Neues Feature	HADOOP-18850: S3A: Aktiviert serverseitige Dual-Layer-Verschlüsselung mit Schlüsseln AWS KMS
Neues Feature	Support S3 Glacier beim Lesen wiederhergestellter Objekte
Verbesserung	Dynamische Skalierung der maximalen S3A-Verbindungen basierend auf den Kernen bei der Ausführung von Spark-Jobs
Verbesserung	HADOOP-18797 : Support gleichzeitige Schreibvorgänge mit S3A Magic Committer
Verbesserung	HADOOP-19047 : Support das InMemory Tracking von S3A Magic Commits
Verbesserung	Fügen Sie die Methoden „Exportieren“, „Veröffentlichen“ und „AbortPublish“ in MagicV2 Committer hinzu
Verbesserung	Implementieren Sie in S3A ReadFullyIntoBuffers FileSystem API
Fehlerbehebung	HADOOP-18793 : S3A bereinigt das StagingCommitter Staging-Uploads-Verzeichnis nicht
Upgrade	HADOOP-19008 : S3A: Aktualisiere die AWS-SDK-Version auf 2.21.41

Typ	Beschreibung
Upgrade	HADOOP-18613 ZooKeeper : Upgrade auf Version 3.8.3

Amazon EMR 7.2.0 — Hadoop-Funktionen

In der folgenden Liste finden Sie neue Hadoop-Funktionen in Amazon EMR 7.2.0.

- [S3-Gletscherunterstützung zum Lesen wiederhergestellter Objekte](#) — Mit Amazon EMR Version 7.2.0 und höher können Sie wiederhergestellte Glacier-Objekte vom S3-Speicherort der Tabelle mit dem S3A Protokoll lesen. Bei diesem Lesevorgang werden archivierte Glacier-Dateien ignoriert, wenn sie noch wiederhergestellt werden.
- Support gleichzeitiger Schreibvorgänge mit dem S3A Magic Committer — Amazon EMR 7.2.0 bietet Unterstützung für gleichzeitige Schreibvorgänge mit dem S3A Magic Committer. [Diese Funktion wurde der Hadoop-Community als Teil von -18797 hinzugefügt. OSS HADOOP](#)

Amazon EMR 6.6.0 — Hadoop-Versionshinweise

Amazon EMR 6.6.0 — Hadoop-Änderungen

Typ	Beschreibung
Fehler	Es wurden doppelte Datensätze beim Lesen BZip2 von Textdateien behoben.
Backport	HADOOP-18136 : Überprüfen Sie den <code>FileUtils.unTar()</code> Umgang mit fehlenden .tar-Dateien
Backport	HADOOP-17627 : Backport auf Branch-3.2 -17371, -17621, HADOOP -17625, um Jetty auf 9.4.39 zu aktualisieren HADOOP HADOOP
Backport	HADOOP-17655 : Aktualisiere Jetty auf 9.4.40

Typ	Beschreibung
Backport	HADOOP-17796 : Aktualisiere die Jetty-Version auf 9.4.43
Backport	HADOOP-17661 : mvn versions:set kann pom.xml nicht analysieren
Backport	HADOOP-17236 : Erhöhen Sie Snakeyaml auf 1.26, um das Problem mit -2017-18640 abzuschwächen CVE
Backport	HADOOP-16717 : jLogger Entfernen GenericsUtil isLog Sie die Abhängigkeit von 4 von Log4jLoggerAdapter
Backport	HADOOP-17633 : JSON-Smart auf 2.4.2 und 9.8 erhöhen aufgrund nimbus-jose-jwt CVEs
Backport	HADOOP-17844 JSON: Aktualisieren Sie Smart auf 2.4.7
Backport	HADOOP-17972 : Backport HADOOP -17683 (Update commons-io auf 2.8.0) für Branch-3.2
Backport	HADOOP-16555 : Aktualisiere Commons-Compress auf Version 1.19
Backport	HADOOP-17370: Aktualisiere Commons-Compress auf Version 1.21
Backport	HADOOP-17096 : Korrigiert den Offset des Eingabepuffers ZStandardCompressor
Backport	HADOOP-17112 : Leerzeichen in Pfaden sind nicht erlaubt, wenn Dateien per Committer in S3A gespeichert werden

Typ	Beschreibung
Backport	HADOOP-13500 : Die Iteration des Objekts mit den Konfigurationseigenschaften wird synchronisiert
Backport	HDFS-14099 : Unbekannter Frame-Deskriptor beim Dekomprimieren mehrerer Frames in ZStandardDecompressor
Backport	HDFS-16410 : Unsicheres Xml-Parsen bei OfflineEditsXmlLoader
Backport	HDFS-14498 : LeaseManager kann die Datei, deren Erstellung fehlgeschlagen ist, für immer wiederholen
Backport	HDFS-15290 : während des Starts aktiviert NPE HttpServer NameNode
Backport	HDFS-15293 : Lockern Sie die Bedingung für die Annahme eines FS-Images beim Empfang eines Checkpoints
Backport	HDFS-12979 : sollte nach dem Checkpoint hochgeladen werden StandbyNode FsImage ObserverNode
Backport	YARN-10538 : Fügt Knoten für die Wiederinbetriebnahme zur Liste der aktualisierten Knoten hinzu, die an den AM zurückgegeben wurden
Backport	YARN-10472 : Backport YARN -10314 (YarnClient wirft NoClassDefFoundError für nur schattierte Client-Jars) nach Branch-3.2 WebSocketException

Typ	Beschreibung
Backport	YARN-9968 : Public Localizer wird aus folgenden Gründen beendet NodeManager NullPointerException
Backport	YARN-10651 : ist mit in abgestürzt CapacityScheduler NPE AbstractYarnScheduler.updateNodeResource()
Backport	YARN-9339 : Die Metrik für ausstehende Apps ist falsch, nachdem die App in eine neue Warteschlange verschoben wurde
Backport	YARN-10438 : Behandle Null in containerId ClientRMService#getContainerReport()
Backport	YARN-7266 : ATS 1.5 kann nicht gestartet werden, wenn RollingLevelDb Dateien beschädigt sind oder fehlen
Backport	YARN-9063 : ATS 1.5 kann nicht gestartet werden, wenn RollingLevelDb Dateien beschädigt sind oder fehlen
Backport	YARN-9848 : Stellt YARN -4946 wieder her (RM sollte eine Anwendung nicht so betrachten, COMPLETED als ob sich die Protokollaggregation nicht im Terminalzustand befindet).

Apache HBase

[HBase](#) ist eine nicht relationale, verteilte Open-Source-Datenbank, die im Rahmen des Hadoop-Projekts der Apache Software Foundation entwickelt wurde. HBase läuft auf dem Hadoop Distributed File System (HDFS), um nicht-relationale Datenbankfunktionen für das Hadoop-Ökosystem bereitzustellen. HBase ist in der EMR Amazon-Release-Version 4.6.0 und höher enthalten.

HBase arbeitet nahtlos mit Hadoop zusammen, nutzt das Dateisystem gemeinsam und dient als direkte Eingabe und Ausgabe für das MapReduce Framework und die Ausführungs-Engine. HBase lässt sich auch in Apache Hive integrieren und ermöglicht SQL ähnliche Abfragen über HBase Tabellen, Verknüpfungen mit Hive-basierten Tabellen und Unterstützung für Java Database Connectivity (JDBC). Weitere Informationen über HBase finden Sie unter [Apache HBase](#) und in der [HBase Dokumentation](#) auf der Apache-Website. Ein Beispiel für die Verwendung von HBase mit Hive finden Sie im AWS Big-Data-Blogbeitrag [Combine No SQL and Massively Parallel Analytics using Apache HBase and Apache Hive on Amazon EMR](#).

Mit HBase On Amazon EMR können Sie Ihre HBase Daten auch direkt auf Amazon Simple Storage Service (Amazon S3) sichern und beim Starten eines HBase Clusters aus einem zuvor erstellten Backup wiederherstellen. Amazon EMR bietet zusätzliche Optionen zur Integration mit Amazon S3 für Datenpersistenz und Notfallwiederherstellung.

- **HBase auf Amazon S3** — Mit Amazon EMR Version 5.2.0 und höher können Sie HBase auf Amazon S3 verwenden, um das HBase Stammverzeichnis und die Metadaten eines Clusters direkt in Amazon S3 zu speichern. Anschließend können Sie einen neuen Cluster starten; dieser muss auf den Speicherort im Stammverzeichnis in Amazon S3 verweisen. Nur jeweils ein Cluster kann den HBase Standort in Amazon S3 verwenden, mit Ausnahme eines Read-Replica-Clusters. Weitere Informationen finden Sie unter [HBase auf Amazon S3 \(Amazon S3 S3-Speichermodus\)](#).
- **HBase Read-Replicas** — Amazon EMR Version 5.7.0 und höher HBase auf Amazon S3 unterstützt Read-Replica-Cluster. Ein Read Replica-Cluster bietet schreibgeschützten Zugriff auf die Dateien und Metadaten des Speichers eines primären Clusters, um schreibgeschützte Operationen durchzuführen. Weitere Informationen finden Sie unter [Verwenden eines Lesereplikat-Clusters](#).
- **HBase Snapshots** — Als Alternative zu HBase auf Amazon S3 können Sie mit EMR Version 4.0 und höher Snapshots Ihrer HBase Daten direkt in Amazon S3 erstellen und dann Daten mithilfe der Snapshots wiederherstellen. Weitere Informationen finden Sie unter [Snapshots verwenden](#).

⚠ Important

Für die EMR HBase Amazon-Cluster-Skalierung empfehlen wir nicht, [verwaltete Skalierung](#) oder [Skalierung mit benutzerdefinierten Richtlinien für HBase Cluster](#) zu verwenden.

In der folgenden Tabelle sind die Version von HBase aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. HBase

Informationen zur Version der Komponenten, mit denen HBase in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

HBaseVersionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	HBaseAusführung	Mit HBase installierte Komponenten
emr-7.2.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

In der folgenden Tabelle sind die Version von HBase aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. HBase

Informationen zur Version der Komponenten, mit denen HBase in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 6.15.0](#).

HBaseVersionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	HBaseAusführung	Mit HBase installierte Komponenten
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-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

Note

Apache HBase HBCK2 ist ein separates Betriebssystem zur Reparatur von HBase Regionen und Systemtabellen. In EMR Amazon-Version 6.1.0 und höher wird die Datei `hbase-hbck2.jar` `/usr/lib/hbase-operator-tools/` auf dem primären Knoten bereitgestellt. Weitere Informationen zur Erstellung und Verwendung des Tools finden Sie unter [HBaseHBCK2](#).

In der folgenden Tabelle sind die Version von HBase aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. HBase

Informationen zur Version der Komponenten, mit denen HBase in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

HBaseVersionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	HBaseAusführung	Mit HBase installierte Komponenten
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

Themen

- [Einen Cluster erstellen mit HBase](#)
- [HBase auf Amazon S3 \(Amazon S3 S3-Speichermodus\)](#)
- [Write-Ahead-Logs \(WAL\) für Amazon EMR](#)
- [Verwenden der HBase Shell](#)
- [Greifen Sie mit Hive auf HBase Tabellen zu](#)
- [Schnappschüsse verwenden HBase](#)
- [Konfigurieren HBase](#)

- [Sehen Sie sich die HBase Benutzeroberfläche an](#)
- [Anzeige von HBase-Protokolldateien](#)
- [Monitor HBase mit Ganglia](#)
- [Migration von früheren Versionen HBase](#)
- [HBase Versionsverlauf](#)

Einen Cluster erstellen mit HBase

Die Verfahren in diesem Abschnitt behandeln die Grundlagen des Startens eines Clusters mithilfe von AWS Management Console und AWS CLI. Ausführliche Informationen zur Planung, Konfiguration und Inbetriebnahme von EMR Amazon-Clustern finden Sie unter [Planen und Konfigurieren von Clustern](#) im Amazon EMR Management Guide.

Einen Cluster HBase mithilfe der Konsole erstellen

Schnelle Schritte zum Starten von Clustern mit der Konsole finden Sie unter [Getting started with Amazon EMR](#) im Amazon EMR Management Guide.

Um einen Cluster zu starten, der über die Konsole HBase installiert ist

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie unter Software Configuration (Softwarekonfiguration) die Amazon Release Version 4.6.0 oder höher aus (wir empfehlen die neueste Version). Wählen Sie nach HBaseBedarf weitere Anwendungen aus.
4. Wählen Sie bei EMR Amazon-Version 5.2.0 und höher unter HBaseSpeichereinstellungen HDFSoder S3 aus. Weitere Informationen finden Sie unter [HBaseauf Amazon S3 \(Amazon S3 S3-Speichermodus\)](#).
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Erstellen eines Clusters HBase mit der AWS CLI

Verwenden Sie den folgenden Befehl, um einen Cluster mit HBase installiertem zu erstellen:

```
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

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Wenn Sie HBase auf Amazon S3 verwenden, geben Sie die `--configurations` Option mit einem Verweis auf ein JSON Konfigurationsobjekt an. Das Konfigurationsobjekt muss eine `hbase-site` Klassifizierung enthalten, die den Speicherort in Amazon S3 angibt, an dem HBase Daten mithilfe der `hbase.rootdir` Eigenschaft gespeichert werden. Es muss außerdem eine `hbase`-Klassifizierung enthalten, die `s3` über die Eigenschaft `hbase.emr.storageMode` angibt. Das folgende Beispiel zeigt einen JSON Ausschnitt mit diesen Konfigurationseinstellungen.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3"  
    }  
  }  
]
```

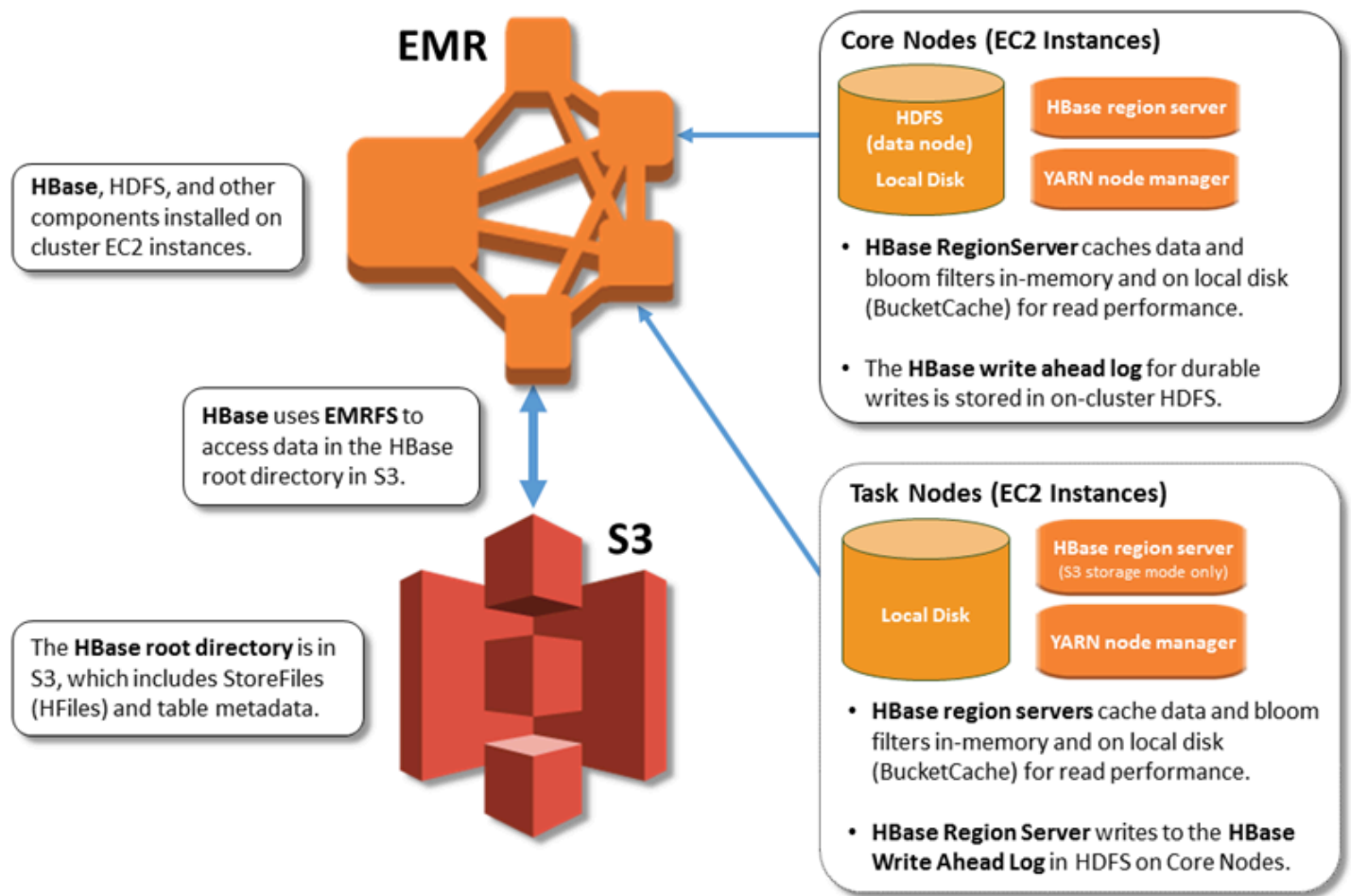
Weitere Informationen zu HBase Amazon S3 finden Sie unter [HBase auf Amazon S3 \(Amazon S3 S3-Speichermodus\)](#). Weitere Informationen zu Klassifizierungen finden Sie unter [Anwendungen konfigurieren](#).

HBase auf Amazon S3 (Amazon S3 S3-Speichermodus)

Wenn Sie HBase auf Amazon EMR Version 5.2.0 oder höher ausführen, können Sie die Aktivierung HBase auf Amazon S3 durchführen, was die folgenden Vorteile bietet:

- Das HBase Stammverzeichnis wird in Amazon S3 gespeichert, einschließlich HBase Speicherdateien und Tabellenmetadaten. Diese Daten sind außerhalb des Clusters persistent und in allen Amazon EC2 Availability Zones verfügbar, und Sie müssen sie nicht mithilfe von Snapshots oder anderen Methoden wiederherstellen.
- Wenn Sie Dateien in Amazon S3 speichern, können Sie die Größe Ihres EMR Amazon-Clusters an Ihre Rechenanforderungen und nicht an die Datenanforderungen anpassen, und zwar mit 3-facher Replikation HDFS.
- Mit EMR Amazon-Version 5.7.0 oder höher können Sie einen Read-Replica-Cluster einrichten, mit dem Sie schreibgeschützte Kopien von Daten in Amazon S3 verwalten können. Sie können auf die Daten vom Read Replica-Cluster zugreifen, um gleichzeitig Lesevorgänge auszuführen und wenn der primäre Cluster nicht mehr verfügbar ist.
- In EMR Amazon-Version 6.2.0 und höher verwendet persistentes HFile Tracking eine HBase Systemtabelle, die aufgerufen wird `hbase:storefile`, um die für Lesevorgänge verwendeten HFile Pfade direkt zu verfolgen. Dieses Feature ist standardmäßig aktiviert und erfordert keine manuelle Migration.

Die folgende Abbildung zeigt die HBase Komponenten, die für HBase Amazon S3 relevant sind.



Aktivierung HBase auf Amazon S3

Sie können die Aktivierung HBase auf Amazon S3 mit der EMR Amazon-Konsole, AWS CLI, dem Amazon EMR API oder Amazon EMR durchzuführen. Die Konfiguration wird während der Cluster-Erstellung als Option angeboten. Wenn Sie die Konsole verwenden, wählen Sie die Einstellung über **Advanced options (Erweiterte Optionen)** aus. Wenn Sie das verwenden AWS CLI, verwenden Sie die `--configurations` Option, um ein JSON Konfigurationsobjekt bereitzustellen. Die Eigenschaften des Konfigurationsobjekts geben den Speichermodus und das Stammverzeichnis in Amazon S3 an. Der von Ihnen angegebene Amazon S3 S3-Standort sollte sich in derselben Region wie Ihr EMR Amazon-Cluster befinden. In Amazon S3 kann jeweils nur ein aktiver Cluster dasselbe HBase Stammverzeichnis verwenden. Anweisungen zur Konsole und ein detailliertes Beispiel zum Erstellen eines Clusters mithilfe von finden Sie unter AWS CLI. [Einen Cluster erstellen mit HBase](#) Ein Beispiel für ein Konfigurationsobjekt wird im folgenden JSON Codeausschnitt gezeigt.

```
{
```

```
"Classification": "hbase-site",
"Properties": {
  "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3"
  }
}
```

Note

Wenn Sie einen Amazon S3-Bucket als `rootdir` für verwenden HBase, müssen Sie am Ende des Amazon S3 URI einen Schrägstrich hinzufügen. Zum Beispiel müssen Sie statt `"hbase.rootdir: s3://my-bucket/"` `"hbase.rootdir: s3://my-bucket"` verwenden, um Probleme zu vermeiden.

Verwenden eines Lesereplikat-Clusters

Nachdem Sie HBase auf Amazon S3 einen primären Cluster eingerichtet haben, können Sie einen Read-Replica-Cluster erstellen und konfigurieren, der schreibgeschützten Zugriff auf dieselben Daten wie der primäre Cluster bietet. Dies ist hilfreich, wenn Sie gleichzeitigen Zugriff auf Abfragedaten brauchen oder einen unterbrechungsfreien Zugriff benötigen, wenn der primäre Cluster nicht mehr verfügbar ist. Die Read-Replica-Funktion ist mit EMR Amazon-Version 5.7.0 und höher verfügbar.

Der primäre Cluster sowie der Read Replica-Cluster werden auf die gleiche Weise eingerichtet, mit einem wichtigen Unterschied. Beide verweisen auf denselben `hbase.rootdir`-Speicherort. Die hbase-Klassifizierung für den Read Replica-Cluster enthält jedoch die Eigenschaft `"hbase.emr.readreplica.enabled": "true"`.

Ausgehend von der JSON Klassifizierung für den primären Cluster, wie weiter oben in diesem Thema beschrieben, sieht die Konfiguration für einen Read-Replica-Cluster beispielsweise wie folgt aus:

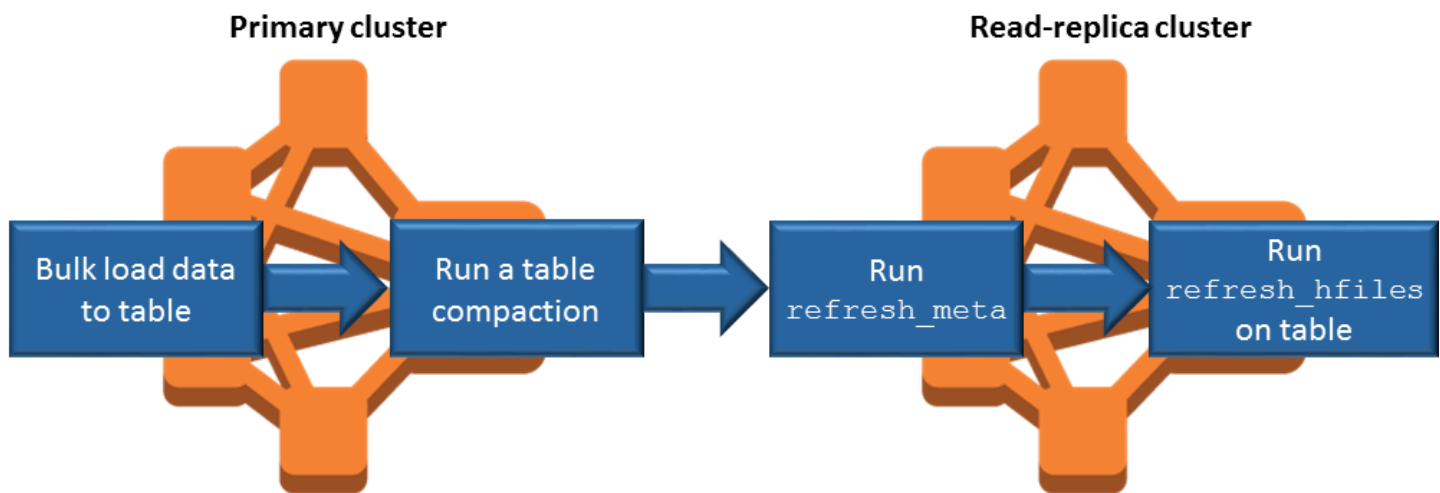
```
{
  "Classification": "hbase-site",
  "Properties": {
```

```
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.readreplica.enabled": "true"
    }
  }
}
```

Synchronisieren des Lesereplikats beim Hinzufügen von Daten

Da die Read-Replica Metadaten verwendet HBase StoreFiles, die der primäre Cluster in Amazon S3 schreibt, ist die Read-Replica nur so aktuell wie der Amazon S3 S3-Datenspeicher. Die folgenden Anleitungen können Ihnen dabei helfen, die Zeitverzögerung beim Schreiben von Daten zwischen dem primären Cluster und dem Read Replica zu minimieren.

- Sie sollten nach Möglichkeit Masseneinladevorgänge für den primären Cluster durchführen. [Weitere Informationen finden Sie unter Bulk Loading in der Apache-Dokumentation.](#) HBase
- Möglichst bald nach dem Hinzufügen der Daten sollte eine Auslagerung erfolgen, bei der Speicherdateien in Amazon S3 geschrieben werden. Führen Sie entweder eine manuelle Auslagerung durch oder optimieren Sie die Auslagerungseinstellungen, um Verzögerungen zu minimieren.
- Wenn Verdichtungen möglicherweise automatisch ausgeführt werden, führen Sie eine manuelle Verdichtung durch, um Inkonsistenzen bei der Auslösung von Verdichtungen zu vermeiden.
- Führen Sie auf dem Read-Replica-Cluster den Befehl aus, wenn sich Metadaten geändert haben, z. B. wenn HBase Regionen aufgeteilt oder komprimiert werden oder wenn Tabellen hinzugefügt oder entfernt werden. `refresh_meta`
- Führen Sie den Befehl `refresh_hfiles` auf dem Read Replica-Cluster aus, wenn Datensätze zu einer Tabelle hinzugefügt oder in einer Tabelle geändert werden.



HFile Persistentes Tracking

Persistentes HFile Tracking verwendet eine HBase Systemtabelle, die aufgerufen wird `hbase:storefile`, um die für Lesevorgänge verwendeten HFile Pfade direkt nachzuverfolgen. Neue HFile Pfade werden der Tabelle hinzugefügt, wenn zusätzliche Daten hinzugefügt werden HBase. Dadurch werden Umbenennungsvorgänge als Festschreibungsmechanismus bei kritischen HBase Schreibpfadvorgängen entfernt und die Wiederherstellungszeit beim Öffnen einer HBase Region verbessert, indem aus der `hbase:storefile` Systemtabelle statt aus der Verzeichnisliste des Dateisystems gelesen wird. Diese Funktion ist in EMR Amazon-Version 6.2.0 und höher standardmäßig aktiviert und erfordert keine manuellen Migrationsschritte.

Note

Die persistente HFile Nachverfolgung mithilfe der HBase Storefile-Systemtabelle unterstützt die Funktion zur HBase Regionsreplikation nicht. Weitere Informationen zur HBase Regionsreplikation finden Sie unter [Timeline-consistent High Available Reads](#).

Persistentes Tracking deaktivieren HFile

Persistentes HFile Tracking ist ab EMR Amazon-Version 6.2.0 standardmäßig aktiviert. Um persistentes HFile Tracking zu deaktivieren, geben Sie beim Starten eines Clusters die folgende Konfigurationsüberschreibung an:

```
{
  "Classification": "hbase-site",
```



```
"Properties": {
  "hbase.storefile.tracking.persist.enabled":"false",

"hbase.hstore.engine.class":"org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
}
}
```

Note

Bei der Neukonfiguration des EMR Amazon-Clusters müssen alle Instanzgruppen aktualisiert werden.

Manuelles Synchronisieren der Storefile-Tabelle

Die Storefile-Tabelle wird auf dem neuesten Stand gehalten, sobald neue erstellt HFiles werden. Wenn die Storefile-Tabelle jedoch aus irgendeinem Grund nicht mehr mit den Datendateien synchron ist, können die folgenden Befehle verwendet werden, um die Daten manuell zu synchronisieren:

Synchronisieren der Storefile-Tabelle in einer Online-Region:

```
hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>
```

Synchronisieren der Storefile-Tabelle in einer Offline-Region:

- Entfernen Sie die Storefile-Tabelle znode.

```
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[<tableName>, hbase:namespace]
# The TableName exists in the list
echo "delete /hbase/storefile/loaded/<tableName>" | sudo -u hbase hbase zkcli
# Delete the Table ZNode
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[hbase:namespace]
```

- Weisen Sie die Region zu (in 'hbase shell' ausführen).

```
hbase cli> assign '<region name>'
```

- Wenn die Zuweisung fehlschlägt.

```
hbase cli> disable '<table name>'
```

```
hbase cli> enable '<table name>'
```

Skalierung der Storefile-Tabelle

Die Storefile-Tabelle ist standardmäßig in vier Regionen aufgeteilt. Wenn die Storefile-Tabelle immer noch unter starker Schreiblast steht, kann die Tabelle manuell weiter aufgeteilt werden.

Um eine bestimmte Hot-Region aufzuteilen, verwenden Sie den folgenden Befehl (in 'hbase shell' ausführen).

```
hbase cli> split '<region name>'
```

Um die Tabelle zu teilen, verwenden Sie den folgenden Befehl (in der „hbase-Shell“ ausführen).

```
hbase cli> split 'hbase:storefile'
```

Betriebliche Überlegungen

HBaseRegionalservers werden verwendet BlockCache, um gelesene Daten im Arbeitsspeicher und BucketCache um gelesene Daten auf der lokalen Festplatte zu speichern. Darüber hinaus speichern Regionalserver Datenschreibvorgänge im Arbeitsspeicher und verwenden MemStore Write-Ahead-Protokolle, um Datenschreibvorgänge zu speichern, HDFS bevor die Daten in Amazon S3 geschrieben werden. HBase StoreFiles Die Leseleistung Ihres Clusters hängt davon ab, wie oft ein Datensatz vom In-Memory-Cache oder Cache auf dem Datenträger abgerufen werden kann. Ein Cache-Fehler führt dazu, dass der Datensatz aus Amazon S3 gelesen wird, was eine deutlich höhere Latenz und eine höhere Standardabweichung aufweist als das Lesen von HDFS. StoreFile Darüber hinaus sind die maximalen Anforderungsraten für Amazon S3 geringer als das, was vom lokalen Cache erreicht werden kann. Also kann die Zwischenspeicherung von Daten für leseintensive Workloads sehr wichtig sein. Weitere Informationen zu Amazon-S3-Leistung finden Sie unter [Optimieren der Leistung](#) im Benutzerhandbuch zu Amazon Simple Storage Service.

Um die Leistung zu verbessern, empfehlen wir, so viel wie möglich von Ihrem Datensatz im EC2 Instance-Speicher zwischenspeichern. Da der den EC2 Instance-Speicher des Regionsservers BucketCache verwendet, können Sie einen EC2 Instance-Typ mit ausreichend Instance-Speicher wählen und EBS Amazon-Speicher hinzufügen, um der erforderlichen Cache-Größe gerecht zu werden. Mithilfe der `hbase.bucketcache.size` Eigenschaft können Sie auch die BucketCache Größe der angehängten Instance-Speicher und EBS Volumes erhöhen. Die Standardeinstellung lautet 8 192 MB.

Bei Schreibvorgängen können die Häufigkeit von MemStore Leerungen und die Anzahl der bei kleineren und größeren Komprimierungen StoreFiles vorhandenen Leerungen erheblich zu einer Verlängerung der Antwortzeiten der regionalen Server beitragen. Um eine optimale Leistung zu erzielen, sollten Sie erwägen, die Größe des MemStore Flush- und HRegion Blockmultiplikators zu erhöhen, wodurch die Zeit zwischen größeren Verdichtungen verlängert wird, aber auch die Verzögerung bei der Konsistenz zunimmt, wenn Sie eine Read-Replica verwenden. In einigen Fällen erzielen Sie möglicherweise eine bessere Leistung, wenn Sie größere Dateiblöcke (aber weniger als 5 GB) verwenden, um die mehrteilige Upload-Funktionalität von Amazon S3 auszulösen. EMRFS EMR Die Blockgröße von Amazon ist standardmäßig 128 MB. Weitere Informationen finden Sie unter [HDFS Konfiguration](#). Während dem Benchmarking der Leistung durch Bereinigungen und Verdichtungen sind selten Kunden anzutreffen, die die 1 GB-Blockgröße überschreiten. Darüber hinaus funktionieren HBase Komprimierungen und Regionalserver optimal, wenn weniger StoreFiles komprimiert werden müssen.

Es kann recht lange dauern, bis Tabellen in Amazon S3 gelöscht sind, da große Verzeichnisse umbenannt werden müssen. Erwägen Sie, Tabellen zu deaktivieren, anstatt sie zu entfernen.

Es gibt einen HBase saubereren Prozess, der alte WAL Dateien und Speicherdateien bereinigt. Mit der EMR Amazon-Version 5.17.0 und höher ist der Cleaner global aktiviert, und die folgenden Konfigurationseigenschaften können verwendet werden, um das Cleaner-Verhalten zu steuern.

Konfigurationseigenschaft	Standardwert	Beschreibung
<code>hbase.regionserver.hfilecleaner.large.thread.count</code>	1	Die Anzahl der Threads, die der Säuberung zugewiesen wurden, ist groß abgelaufen. HFiles
<code>hbase.regionserver.hfilecleaner.small.thread.count</code>	1	Die Anzahl der zum Säubern zugewiesenen Threads ist gering abgelaufen HFiles.
<code>hbase.cleaner.scan.dir.concurrent.size</code>	Setzen Sie diesen Wert auf ein Viertel aller verfügbaren Kerne.	Die Anzahl der Threads zum Scannen der oldWALS Verzeichnisse.

Konfigurationseigenschaft	Standardwert	Beschreibung
<code>hbase.oldwals.cleaner.thread.size</code>	2	Die Anzahl der Threads, die WALs unter dem oldWALS Verzeichnis gesäubert werden sollen.

Bei Amazon EMR 5.17.0 und früheren Versionen kann der Cleaner-Vorgang die Abfrageleistung beeinträchtigen, wenn hohe Workloads ausgeführt werden. Wir empfehlen daher, den Cleaner-Vorgang nur außerhalb der Spitzenzeiten zu aktivieren. Der Cleaner hat die folgenden Shell-Befehle:

HBase

- `cleaner_chore_enabled` fragt ab, ob die Bereinigungsfunktion aktiviert ist.
- `cleaner_chore_run` führt die Bereinigungsfunktion manuell aus, um Dateien zu entfernen.
- `cleaner_chore_switch` aktiviert oder deaktiviert die Bereinigungsfunktion und gibt den vorherigen Zustand der Bereinigungsfunktion zurück. Durch `cleaner_chore_switch true` wird die Bereinigungsfunktion beispielsweise aktiviert.

Eigenschaften für die Leistungsoptimierung HBase auf Amazon S3

Die folgenden Parameter können angepasst werden, um die Leistung Ihres Workloads zu optimieren, wenn Sie ihn HBase auf Amazon S3 verwenden.

Konfigurationseigenschaft	Standardwert	Beschreibung
<code>hbase.bucketcache.size</code>	8,192	Die Menge an Festplatte, die auf Regionsservern reserviert ist, die Amazon EC2 Instance speichert, und die EBS Volumes für die BucketCache Speicherung. Die Einstellung gilt für alle Regionsserver-Instances. Größere BucketCache Größen entsprechen in

Konfigurationseigenschaft	Standardwert	Beschreibung
		der Regel einer verbesserten Leistung
<code>hbase.hregion.memstore.flush.size</code>	134217728	Das Datenlimit in Byte, bei dem eine MemStore-Auslagerung nach Amazon S3 ausgelöst wird.
<code>hbase.hregion.memstore.block.multiplier</code>	4	Ein Multiplikator, der die MemStore Obergrenze bestimmt, ab der Updates blockiert werden. Wenn die MemStore <code>hbase.hregion.memstore.flush.size</code> Überschreitungen mit diesem Wert multipliziert werden, werden Updates blockiert. MemStore Aktualisierungen können durch Leeren und Komprimieren entsperrt werden.
<code>hbase.hstore.blockingStoreFiles</code>	10	Die maximale Anzahl StoreFiles davon kann in einem Store existieren, bevor Updates blockiert werden.
<code>hbase.hregion.max.filesize</code>	10737418240	Die maximale Größe einer Region, bevor die Region aufgeteilt wird.

Herunterfahren und Wiederherstellen eines Clusters ohne Datenverlust

Um einen EMR Amazon-Cluster herunterzufahren, ohne Daten zu verlieren, die nicht in Amazon S3 geschrieben wurden, sollten Sie Ihren MemStore Cache auf Amazon S3 leeren, um neue Speicherdateien zu schreiben. Zunächst müssen Sie alle Tabellen deaktivieren. Die folgende

Schrittkonfiguration kann verwendet werden, wenn Sie dem Cluster einen Schritt hinzufügen. Weitere Informationen finden Sie im Amazon EMR Management Guide unter [Arbeiten mit Schritten mithilfe der AWS CLI UND-Konsole](#).

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/bin/disable_all_tables.sh"]
```

Alternativ können Sie auch den folgenden Bash-Befehl direkt ausführen.

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

Nachdem Sie alle Tabellen deaktiviert haben, leeren Sie die `hbase:meta` Tabelle mit der HBase Shell und dem folgenden Befehl.

```
flush 'hbase:meta'
```

Anschließend können Sie ein auf dem EMR Amazon-Cluster bereitgestelltes Shell-Skript ausführen, um den MemStore Cache zu leeren. Sie können es als einen Schritt hinzufügen oder direkt auf dem Cluster AWS CLI ausführen. Das Skript deaktiviert alle HBase Tabellen, was dazu führt, dass der Server MemStore in jeder Region zu Amazon S3 flusht. Wenn das Skript erfolgreich abgeschlossen wurde, verbleiben die Daten in Amazon S3 und der Cluster kann beendet werden.


Um einen Cluster mit denselben HBase Daten neu zu starten, geben Sie denselben Amazon S3 S3-Standort wie der vorherige Cluster entweder in der AWS Management Console oder mithilfe der `hbase.rootdir` Konfigurationseigenschaft an.

Write-Ahead-Logs (WAL) für Amazon EMR

Mit Amazon EMR 6.15 und höher können Sie Ihre Apache HBase Write-Ahead-Logs (WAL) auf Amazon schreiben. EMR WAL Bei niedrigeren EMR Amazon-Versionen wird beim Erstellen eines Clusters mit der Option HBase on Amazon S3 die einzige HBase Apache-Komponente auf der lokalen Festplatte für Cluster gespeichert, und Sie können andere Komponenten wie das Stammverzeichnis, store files (HFiles), Tabellenmetadaten und Daten auf Amazon S3 speichern. WAL

Sie können Amazon verwenden EMRWAL, um Daten wiederherzustellen, die nicht auf Amazon S3 übertragen wurden. Um Ihre HBase Cluster vollständig zu sichern, entscheiden Sie sich für die Nutzung des EMR WAL Amazon-Service. `RegionServer` schreibt hinter den Kulissen Ihre HBase Write-Ahead-Logs (WAL) in das WAL für Amazon. EMR

Für den Fall, dass Ihr Cluster oder die AZ fehlerhaft oder nicht verfügbar sind, können Sie einen neuen Cluster erstellen, ihn auf dasselbe S3-Stammverzeichnis und denselben EMR WAL Amazon-Workspace verweisen und die Daten WAL innerhalb weniger Minuten automatisch wiederherstellen. Weitere Informationen finden Sie unter [Von Amazon wiederherstellen EMR WAL](#).

 Note

Amazon EMR bewahrt Ihr Write-Ahead-Protokoll und die zugehörigen Daten ab dem Zeitpunkt, an dem Sie Ihren Cluster erstellen, 30 Tage lang auf. Nach 30 Tagen löscht Amazon EMR automatisch Ihren Amazon EMR WAL und seine Daten. Wenn Sie jedoch einen neuen WAL -fähigen Cluster aus demselben S3-Stammverzeichnis starten, können Sie die Nutzung Ihres Clusters ab dem Startzeitpunkt des neuen Clusters WAL für 30 Tage verlängern. Amazon EMR bereinigt nach Ablauf der ersten 30 Tage weiterhin alle WAL Daten aus dem ersten Cluster. Weitere Informationen finden Sie unter [Von Amazon wiederherstellen EMR WAL](#).

In den folgenden Abschnitten wird beschrieben, wie Sie Amazon EMR WAL mit Ihrem HBase -fähigen EMR Cluster einrichten und verwenden.

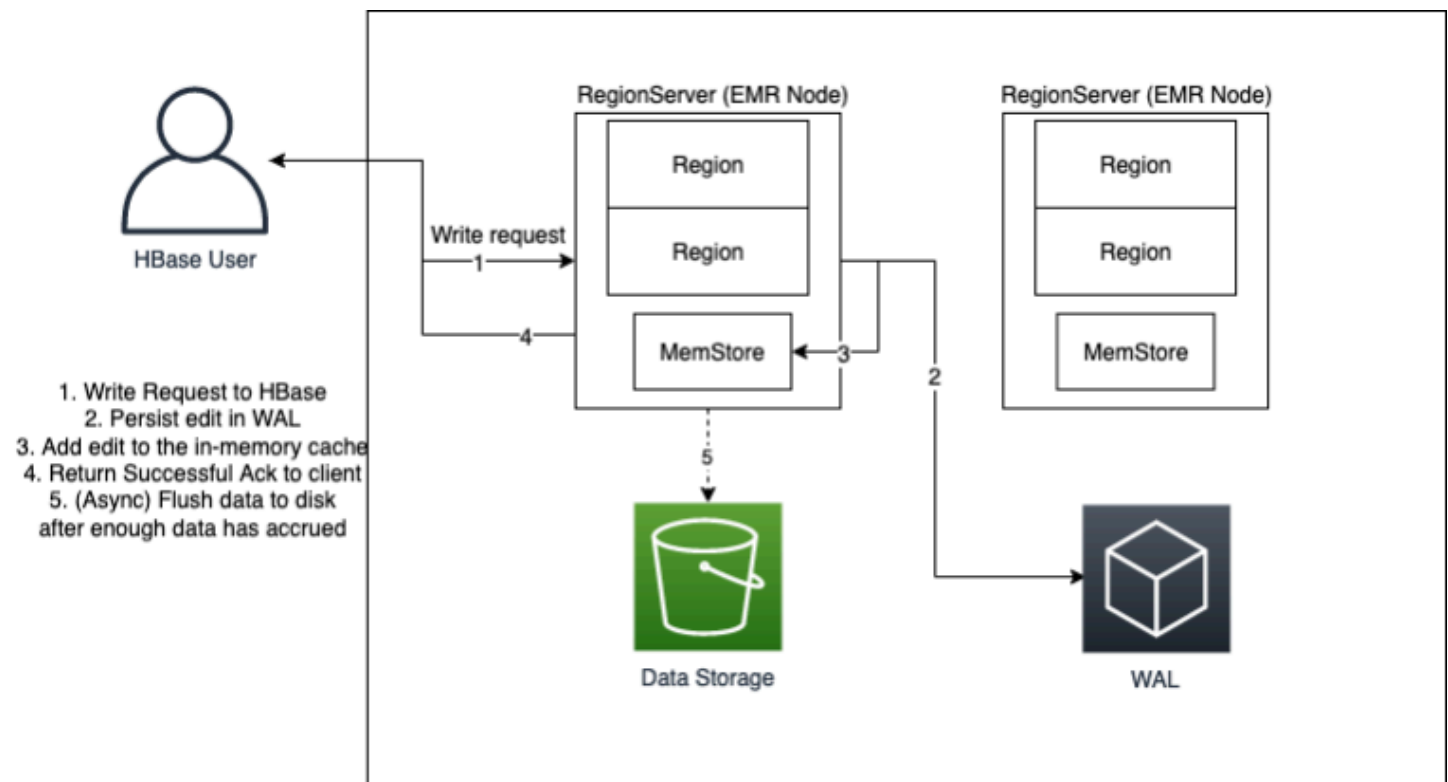
Themen

- [EMRWALAmazon-Arbeitsbereiche](#)
- [Erforderliche Berechtigungen für Amazon EMR WAL](#)
- [Amazon aktivieren EMR WAL](#)
- [Von Amazon wiederherstellen EMR WAL](#)
- [Verwenden von Sicherheitskonfigurationen mit Amazon EMR WAL](#)
- [Greifen Sie auf Amazon zu EMR WAL über AWS PrivateLink](#)
- [Die EMR WAL Preise und Kennzahlen von Amazon verstehen](#)
- [Arbeitsbereiche taggen WAL](#)
- [Überlegungen und Regionen für Amazon EMR WAL](#)
- [Amazon EMR WAL \(EMRWAL\) CLI Referenz](#)

EMRWAL Amazon-Arbeitsbereiche

Amazon EMR WAL fügt das Konzept der WAL Arbeitsbereiche hinzu. Ein WALWorkspace ist ein logischer Container von. WALs Jedes Write-Ahead-Protokoll in Amazon EMR WAL ist von einem Workspace gekapselt. WAL Ein EMR Cluster schreibt WALs in genau einen WAL Workspace, den Sie beim Cluster-Start konfigurieren, oder in den, defaultWALworkspace wenn Sie keinen Workspace angeben. WALWorkspaces haben nichts mit bestehenden HBase Terminologien wie Namespaces zu tun.

Sie können WAL Workspaces verwenden, um EMR WAL IAM Amazon-Berechtigungen so einzuschränken, dass sie nur die Workspaces umfassen, auf die der Cluster zugreifen muss. Sie können Ihren WAL Workspace auch für die Tag-basierte Zugriffskontrolle taggen. Weitere Informationen zum Taggen findest du unter. [Arbeitsbereiche taggen WAL](#)



Erforderliche Berechtigungen für Amazon EMR WAL

Damit Ihr Cluster eine Verbindung zu Amazon herstellen kann EMRWAL, benötigt das Instance-Profil für den Cluster bestimmte IAM Berechtigungen:

- Amazon EMR WAL verwendet die [AWSServiceRoleForEMRWAL](#) serviceverknüpfte Rolle, um einen Cluster-Status abzurufen. Amazon erstellt diese serviceverknüpfte Rolle EMR automatisch, wenn

Sie einen WAL Workspace erstellen, oder HBase erstellt die serviceverknüpfte Rolle, wenn Sie einen Workspace für Amazon konfigurieren EMR WAL und die serviceverknüpfte Rolle noch nicht existiert.

Bevor Sie Amazon EMR WAL für einen Cluster aktivieren können, müssen Sie die Berechtigungen so konfigurieren, dass die `AWSServiceRoleForEMRWAL` serviceverknüpfte Rolle automatisch erstellt werden kann. Weitere Informationen und eine Beispielanweisung, mit der diese Funktion hinzugefügt wird, finden Sie unter [Verwenden von serviceverknüpften Rollen für die Write-Ahead-Protokollierung](#).

- Da Amazon HBase Write Ahead Log (WAL) EMR WAL verwendet, müssen Ihre Cluster dies tun HBaseWAL. Im Folgenden sind die IAM Mindestberechtigungen aufgeführt, die Sie für die Ausführung benötigenHBase. Fügen Sie diese der Berechtigungsrichtlinie für Ihr Instanzprofil hinzu:

```
emrwal:DeleteWal
emrwal:CreateWal
emrwal:CreateWorkspace
emrwal:AppendEdit
emrwal:ReplayEdits
emrwal:GetCurrentWalTime
emrwal:CompleteWalFlush
```

Note

Wenn Sie die Berechtigungen für Amazon nur EMR WAL auf den Mindestsatz beschränken, verfügen einige [EMRWALCLI](#) Befehle nicht über die erforderlichen Berechtigungen zur Ausführung.

Amazon aktivieren EMR WAL

Verwenden Sie die folgenden Schritte, um das Schreiben in den Amazon zu aktivieren EMRWAL, wenn Sie einen Cluster mit dem erstellen AWS Command Line Interface.

Note

Sie können Amazon nicht EMR WAL für einen Cluster aktivieren, der bereits läuft, und Sie können nicht zwei Cluster mit demselben S3-Stammverzeichnis starten. Weitere Informationen finden Sie unter [Überlegungen und Regionen für Amazon EMR WAL](#).

1. Bevor Sie einen EMR WAL Amazon-fähigen Cluster erstellen können, müssen Sie dem Instance-Profil, das Sie mit Ihrem Cluster verwenden möchten, die erforderlichen Berechtigungen hinzufügen. Weitere Informationen finden Sie unter [Erforderliche Berechtigungen für Amazon EMR WAL](#).
2. Erstellen Sie einen Cluster aus dem AWS CLI. Verwenden Sie die `--configurations` Option, um ein JSON Konfigurationsobjekt bereitzustellen, das die `hbase.emr.wal.enabled` Eigenschaft spezifiziert, wie im Beispiel unten gezeigt.
 - Geben Sie den Speichermodus und den Speicherort des Stammverzeichnisses in Amazon S3 an. Der Amazon S3 S3-Standort, den Sie angeben, sollte sich in derselben Region wie Ihr EMR Cluster befinden, aber nur ein aktiver Cluster kann dasselbe HBase Stammverzeichnis in S3 gleichzeitig verwenden.
 - Erstellen Sie Ihren Cluster mit der Konfiguration der Instance-Gruppen. Sie können Amazon nicht EMR WAL mit der Konfiguration der Instance-Flotten verwenden. Weitere Informationen zum Erstellen von Clustern mit Instanzgruppen finden [Sie unter Configure Uniform Instance Groups](#) im Amazon EMR Management Guide.
 - Schritte zur Erstellung eines Clusters in der Konsole und ein ausführliches `create-cluster` Beispiel, das den verwendet AWS CLI, finden Sie unter [Cluster erstellen mit HBase](#).
3. Um die Option WAL für den neuen Cluster zu aktivieren, setzen Sie die `hbase.emr.wal.enabled` Eigenschaft auf `true`. Der folgende Befehl enthält einen JSON Ausschnitt mit einem Beispielkonfigurationsobjekt.

```
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"
    }
  }
]

```

Wenn auf dem neu erstellten Cluster online HBase ist, HBase werden automatisch WAL Daten in den Amazon geschriebenen EMR WAL und der Amazon EMR WAL für Wiederherstellungszwecke verwendet.

Example 1: Einen EMR Cluster erstellen, der Amazon verwendet 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: Einen EMR Cluster mit einem benutzerdefinierten WAL Workspace erstellen

```

[
  {
    "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"  
      }  
    }  
  ]  
]
```

Von Amazon wiederherstellen EMR WAL

Da das Amazon EMR WAL für Ihren ursprünglichen Cluster 30 Tage lang aufbewahrt wird, können Sie es innerhalb dieses Zeitraums von 30 Tagen WAL für einen neu erstellten Cluster wiederherstellen und wiederverwenden. Wenn Sie einen neuen Cluster aus demselben S3-Stammverzeichnis starten, wird die 30-Tage-Uhr ab der Startzeit des neuen Clusters neu gestartet, sofern der vorherige Zeitraum von 30 Tagen nicht abgelaufen ist.

Gehen Sie wie folgt vor, um einen vorhandenen Cluster WAL mit einem neuen Cluster wiederherzustellen. Bei diesem Vorgang wird davon ausgegangen, dass Sie Ihren ursprünglichen Cluster mit EMR WAL aktiviertem Amazon erstellt haben.

1. Erstellen Sie innerhalb von 30 Tagen nach der Erstellung eines Clusters mit WAL aktivierter Option einen neuen Cluster in der selben Cluster AWS-Region wie der ursprüngliche Cluster. Der neue Cluster kann sich in derselben AZ oder in einer anderen AZ innerhalb derselben Region befinden, in der der ursprüngliche Cluster erstellt wurde.

Konfigurieren Sie die Objekteigenschaften, um den Speichermodus und den Speicherort des Stammverzeichnisses in Amazon S3 anzugeben. Der Amazon S3 S3-Standort, den Sie angeben, sollte sich in derselben Region wie Ihr EMR Cluster befinden, aber nur ein aktiver Cluster kann dasselbe HBase Stammverzeichnis in S3 gleichzeitig verwenden.

Konsolenschritte zum Erstellen eines Clusters und ein detailliertes `create-cluster` Beispiel, das den verwendet AWS CLI, finden Sie unter [Cluster erstellen mit HBase](#).

2. Um das bestehende Amazon EMR WAL für den neuen Cluster zu verwenden, setzen Sie die `hbase.emr.wal.enabled` Eigenschaft auf `true`. Der folgende JSON Ausschnitt zeigt ein Beispiel für ein Konfigurationsobjekt.

```
[
```

```

{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
  }
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3",
    "hbase.emr.wal.enabled": "true"
  }
}
]

```

Verwenden von Sicherheitskonfigurationen mit Amazon EMR WAL

Amazon verschlüsselt EMR automatisch sowohl Daten, die zwischen Ihrem Cluster und dem EMR WAL Amazon-Service übertragen werden, als auch die Daten im Ruhezustand in Amazon EMR WAL. Weitere Informationen finden Sie unter [Verschlüsselung im Ruhezustand für Amazon EMR WAL](#). Sie können auch eine Sicherheitskonfiguration verwenden, um Ihre eigenen Schlüssel vom AWS Key Management Service (KMS) -Service mitzubringen und die Daten, die Sie bei Amazon EMR WAL speichern, zu verschlüsseln.

Verwenden Sie eine der folgenden Methoden, um eine Sicherheitskonfiguration auszuwählen, wenn Sie einen Cluster erstellen:

Console

Geben Sie von der AWS Management Console aus die Konfiguration unter Sicherheitskonfiguration und EC2 key pair an.

Security configuration and EC2 key pair - optional [Info](#)

Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.



CLI

Stellen Sie unter den `--security-configuration` Parameter ein AWS CLI, wenn Sie den Befehl [create-cluster](#) verwenden.

Weitere Informationen finden Sie unter [Verschlüsselung im Ruhezustand für Amazon EMR WAL](#) und [Verwenden von Sicherheitskonfigurationen zur Einrichtung der Clustersicherheit](#) im Amazon EMR Management Guide.

Weitere sicherheitsrelevante Informationen WAL finden Sie unter [Verwenden von serviceverknüpften Rollen für](#) die Write-Ahead-Protokollierung.

Greifen Sie auf Amazon zu EMR WAL über AWS PrivateLink

Wenn Sie Ihre Verbindung innerhalb des AWS Netzwerks beibehalten möchten, bietet EMR WAL Amazon AWS PrivateLink Support an. Verwenden Sie zur Einrichtung das AWS Management Console oder AWS Command Line Interface (AWS CLI) AWS PrivateLink, um einen VPC Schnittstellenendpunkt zu erstellen, der eine Verbindung zu Amazon herstellt EMRWAL. Weitere Informationen finden Sie im AWS PrivateLink Leitfaden unter [Zugreifen auf einen AWS Service über einen VPC Schnittstellenendpunkt](#).

Die grundlegenden Schritte sind wie folgt:

1. Verwenden Sie die VPC Amazon-Konsole, um [einen VPC Endpunkt zu erstellen](#). Wählen Sie Endpoints und dann Endpoint erstellen aus.
2. Behalten Sie die Dienstkategorie als AWS Dienste bei.
3. Geben Sie in der Suchleiste für den Bereich Dienste den Dienst mit der Bezeichnung `emrwal`, und wählen Sie ihn aus `com.amazonaws.region.emrwal.prod`.
4. Wählen Sie Ihren Endpunkt aus VPC und speichern Sie ihn. Stellen Sie sicher, dass Sie dem VPC Endpunkt dieselben Sicherheitsgruppen zuordnen, die Sie dem EMR Cluster zuordnen.
5. Wenn Sie möchten, können Sie jetzt private DNS Hostnamen für Ihren neuen Endpunkt aktivieren. Stellen Sie DNS Hostnamen aktivieren und DNSSupport aktivieren auf `true` für Sie VPC ein. Wählen Sie dann Ihre Endpunkt-ID aus, klicken Sie im Menü Aktionen auf VPCEinstellungen bearbeiten und aktivieren Sie private DNS Namen.
 - Die privaten DNS Hostnamen für den Endpunkt folgen dem Format `prod.emrwal.region.amazonaws.com`.
 - Wenn Sie private DNS Hostnamen nicht aktivieren, VPC stellt Amazon Ihnen einen DNS Endpunktnamen im folgenden Format `endpointID.prod.emrwal.region.vpce.amazonaws.com` zur Verfügung.

6. Um Ihren AWS PrivateLink Endpunkt zu verwenden, ändern Sie die `emr.wal.client.endpoint` Konfiguration, wenn Sie Ihren [EMRWAL Amazon-fähigen Cluster](#) erstellen, wie im folgenden Beispiel gezeigt:

```
[
  {
    "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"
    }
  }
]
```

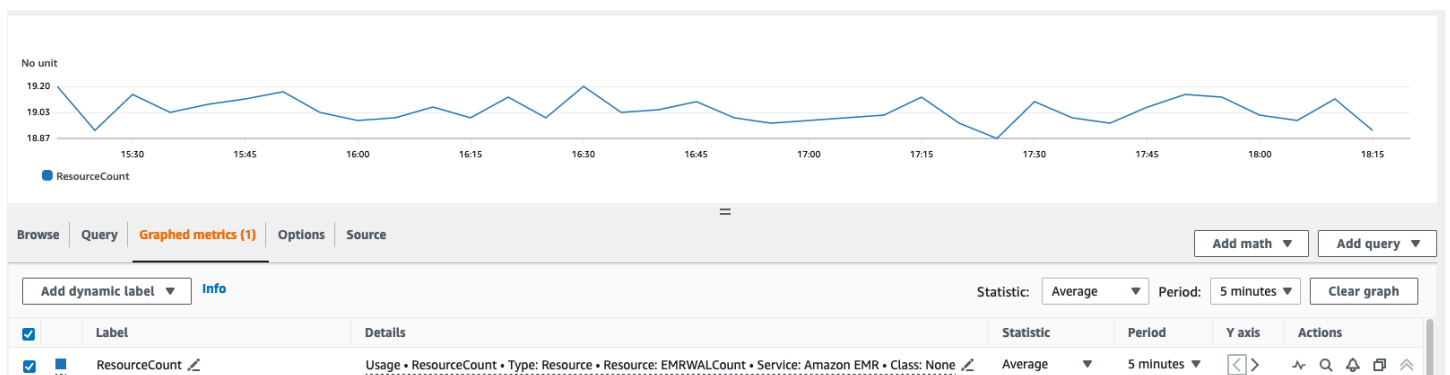
Sie können auch VPCE Richtlinien verwenden, um den Zugriff auf Amazon zu erlauben oder einzuschränken EMR WAL APIs. Weitere Informationen finden Sie im AWS PrivateLink Leitfaden unter [Steuern des Zugriffs auf VPC Endgeräte mithilfe von Endpunktrichtlinien](#).

Die EMR WAL Preise und Kennzahlen von Amazon verstehen

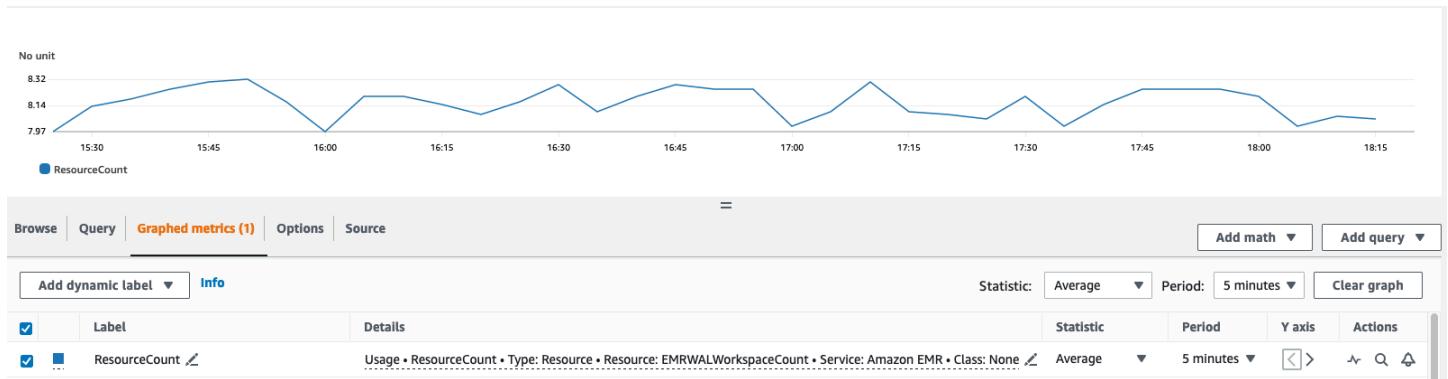
Abrechnungseinheit für Kernfunktionen	Details
EMR-WAL-Read-GiB	API calls to read data from your table are billed as ReadRequestGiB. This includes Get and Scan operations. Reads are charged based on the sizes of the read items. Amazon EMR bills at a minimum of 1 byte. For example, if you read a 1234.12 bytes item, you're charged for 1235 bytes. Reads are aggregated every hour for billing and shown as GiBs.

Abrechnungseinheit für Kernfunktionen	Details
EMR-WAL-Write-GiB	<p>API calls to write data from your table are billed as Write-GiB. This includes Put operations. Writes are charged based on the sizes of the written items. Amazon EMR bills at a minimum of 1 byte. For example, if you write a 1234.12 bytes item, you're charged for 1235 bytes. Writes are aggregated every hour for billing and shown as GiBs.</p>
EMR-WAL-WALHours	<p>Die Anzahl der DatenWALs, die Sie im Service speichern, wird als EMR-WAL-WALHours abgerechnet. Amazon EMR erstellt eine WAL pro HBase Region. Wenn Sie beispielsweise 20 HBase Tabellen einschließlich Systemtabellen erstellen und jede Tabelle zwei HBase Regionen hat, dann benötigen Sie 28.800 WAL Stunden, berechnet als:</p> <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; margin: 10px 0;"> <p style="margin: 0;">20 tables x 2 Regions per table x 1 WAL per Region x 30 days x 24 hours ----- 28,800 EMR-WAL-WALHours</p> </div>

Beispiel: EMRWALCount



Beispiel EMRWALWorkspaceCount:



Arbeitsbereiche taggen WAL

Sie können Tags zu einem Workspace hinzufügen, wenn Sie einen neuen Workspace erstellen, und Sie können Tags aus einem aktiven Workspace für einen laufenden Cluster hinzufügen, entfernen oder auflisten. Sie können die einzelnen Ressourcen im Workspace nicht taggen und Sie können keine vorhandenen Tags aktualisieren. Stattdessen können Sie unerwünschte Tags aus dem Workspace entfernen und sie ersetzen.

Sie können Arbeitsbereiche aus dem EMRWAL CLI taggen. Eine Liste der EMRWAL CLI Befehle zum Markieren von Arbeitsbereichen finden Sie unter [Amazon EMR WAL \(EMRWAL\) CLI Referenz](#)

Die folgende IAM Beispielrichtlinie veranschaulicht ein Szenario, in dem CRUDL Workspace-Operationen nur mit dem richtigen Tagging-Schlüssel `resource_tag_allow_test_key` und -Wert zulässig sind: `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"
          ]
        }
      }
    }
  ]
}
```

```
    ]
  }
}
]
```

Um zu überprüfen, ob das Tag jetzt für Workspace-Operationen erforderlich ist, verwenden Sie, [Amazon EMR WAL \(EMRWAL\) CLI Referenz](#) um den [listTagsForResource](#) Befehl `tagAllowResourceTag` für den Workspace mit dem gewünschten Ressourcen-Tag aufzurufen. Wenn Sie die Bedingung korrekt konfiguriert haben, ist der Befehl erfolgreich.

```
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)
```

Überlegungen und Regionen für Amazon EMR WAL

Überlegungen zu Amazon EMR WAL

In der folgenden Liste werden wichtige Überlegungen und Einschränkungen von Amazon beschrieben EMRWAL:

- Amazon EMR WAL kann mit EMR Amazon-Versionen 6.15.0 und höher verwendet werden.
- Amazon EMR WAL ist ein optionaler, kostenpflichtiger Service. Sie zahlen für das, was Sie nutzen: Lese-, Schreib- und Datenspeicherung. Weitere Informationen finden Sie unter [Die EMR WAL Preise und Kennzahlen von Amazon verstehen](#) und auf der Seite mit den [EMRAmazon-Preisen](#).
- Amazon EMR WAL verwendet HBase Write Ahead Log (WAL). Um Amazon verwenden zu können EMRWAL, müssen Ihre Cluster verwenden HBaseWAL.
- Um Amazon EMR WAL bei der Erstellung eines Clusters zu aktivieren, müssen Sie über die erforderlichen Rollenberechtigungen verfügen. Weitere Informationen finden Sie unter [Verwenden von serviceverknüpften Rollen für die Write-Ahead-Protokollierung](#).
- Sie müssen Amazon aktivieren, EMR WAL wenn Sie den Cluster mit, oder AWS Management Console AWS CLI, erstellenAPI, und Sie müssen die Instance-Gruppenkonfiguration verwenden. Sie können Amazon nicht EMR WAL in einem laufenden Cluster aktivieren, wenn Sie den Cluster nicht mit Amazon erstellt haben EMRWAL. Sie können die `hbase-site` Konfigurationen auch nicht bearbeiten, um Amazon EMR WAL in einem laufenden Cluster zu aktivieren.

- Sie können Amazon nur EMR WAL auf Clustern aktivieren, die Amazon S3 für das Stammverzeichnis verwenden.
- Sie können in Amazon S3 nicht mehrere aktive Cluster im selben HBase Stammverzeichnis haben.
- Sie können Amazon nicht EMR WAL auf Read Replica-Clustern aktivieren.
- WAL wird innerhalb des verwalteten Dienstes über alle Availability Zones hinweg repliziert.
- WAL überdauert den Cluster und bleibt für den nächsten Cluster verfügbar.
- Sie können Amazon nicht EMR WAL während des Starts oder wenn Ihr Cluster betriebsbereit ist (im laufenden Zustand), deaktivieren.
- Informationen zu WAL und Workspace-Limits finden Sie unter [EMRAmazon-Endpunkte und Kontingente](#).

Regionale Verfügbarkeit für Amazon EMR WAL

EMRWAL Der Amazon-Service ist in den folgenden Ländern verfügbar AWS-Regionen:

- `ap-northeast-1` – Asien-Pazifik (Tokio)
- `ap-southeast-1` – Asien-Pazifik (Singapur)
- `ap-south-1` – Asien-Pazifik (Mumbai)
- `ap-southeast-2` – Asien-Pazifik (Sydney)
- `eu-central-1` – Europa (Frankfurt)
- `eu-north-1` – Europa (Stockholm)
- `eu-west-1` – Europa (Irland)
- `us-east-1` – USA Ost (Nord-Virginia)
- `us-east-2` – USA Ost (Ohio)
- `us-west-2` – USA West (Oregon)
- `ap-northeast-2` – Asien-Pazifik (Seoul)

Amazon EMR WAL (EMRWAL) CLI Referenz

Die EMRWAL Befehlszeilenschnittstelle (EMRWAL CLI) ist ein einheitliches Tool zur Verwaltung Ihres Write-Ahead-Logs (WAL) für Amazon. EMR Das wird mit EMR Clustern EMRWAL CLI geliefert, wenn Sie es WAL zum Zeitpunkt der Clustererstellung aktivieren. Weitere Informationen zur Aktivierung finden WAL Sie unter [Write-Ahead-Logs \(WAL\) für Amazon EMR](#).

Das EMRWAL CLI beinhaltet die folgenden Befehle:

Themen

- [createWorkspace](#)
- [deleteWal](#)
- [deleteWorkspace](#)
- [listTagsForResource](#)
- [listWals](#)
- [listWorkspaces](#)
- [tagResource](#)
- [untagResource](#)

createWorkspace

Der createWorkspace Befehl erstellt einen neuen EMR WAL Amazon-Workspace.

Verwendung:

```
emrwal createWorkspace [-tags <tags>] [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Beispiel:

```
aws emr createWorkspace -w examples
```

deleteWal

Der deleteWals Befehl löscht den Amazon EMRWAL, den Sie angeben.

Verwendung:

```
emrwal deleteWal [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-n <walName>] [-N <fullName>] [-R] [-m] [-h]
```

Beispiel:

```
emrwal deleteWal -w examples -p hbasetable -n examplewal
```

deleteWorkspace

Der deleteWorkspace Befehl löscht den von Ihnen angegebenen EMR WAL Amazon-Workspace.

Verwendung:

```
emrwal deleteWorkspace [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Beispiel:

```
emrwal deleteWorkspace -w examples
```

listTagsForResource

Der listTagsForResource Befehl listet alle Schlüssel-Wert-Paar-Tags für den EMR WAL Amazon-Workspace auf, den Sie angeben.

Verwendung:

```
emrwal listTagsForResource -arn {resource-arn} [-e {endpoint}] [-r {Region}] [-h]
```

Beispiel:

```
emrwal listTagsForResource -arn arn:aws:emrwal::1234567891234:workspace/examples
```

listWals

Der listWals Befehl listet alle Amazon EMR WALs in dem von Ihnen angegebenen Workspace auf.

Verwendung:

```
emrwal listWals [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-M {integer}] [-h]
```

Beispiel:

```
emrwal listWals -w examples
```

listWorkspaces

Der `listWorkspaces` Befehl listet alle EMR WAL Amazon-Workspaces auf, die Ihnen zur Verfügung stehen.

Verwendung:

```
emrwal listWorkspaces [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}]  
[-r {Region}] [-M {integer}] [-h]
```

Beispiel:

```
emrwal listWorkspaces
```

tagResource

Der `tagResource` Befehl weist dem von Ihnen angegebenen EMR WAL Amazon-Workspace ein oder mehrere Schlüssel-Wert-Paar-Tags zu.

Verwendung:

```
emrwal tagResource -arn {resource-arn} -tags <tags> [-e {endpoint}] [-r {Region}] [-h]
```

Beispiel:

```
emrwal tagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tags tag_key=tag_value
```

untagResource

Der `untagResource` Befehl hebt die Zuweisung eines oder mehrerer Schlüssel-Wert-Paar-Tags zu dem von Ihnen angegebenen EMR WAL Amazon-Workspace auf.

Verwendung:

```
emrwal untagResource -arn {resource-arn} -tagKeys <tagKeys> [-e {endpoint}] [-  
r {Region}] [-h]
```

Beispiel:

```
emrwal untagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tagKeys tag_key
```

Verwenden der HBase Shell

Nachdem Sie einen HBase Cluster erstellt haben, besteht der nächste Schritt darin, eine Verbindung herzustellen, HBase damit Sie mit dem Lesen und Schreiben von Daten beginnen können (Datenschreibvorgänge werden auf einem Read-Replica-Cluster nicht unterstützt). Sie können die [HBaseShell verwenden, um Befehle zu testen](#).

Um die HBase Shell zu öffnen

1. Wird verwendet SSH, um eine Verbindung zum Hauptserver im HBase Cluster herzustellen. Informationen darüber, wie Sie mithilfe von Amazon Connect zum EMR primären Knoten herstellen können SSH, finden [Sie unter Verbindung zum primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Führen Sie `hbase shell`. Die HBase Shell wird mit einer Aufforderung ähnlich der folgenden geöffnet.

```
hbase(main):001:0>
```

Sie können HBase Shell-Befehle von der Eingabeaufforderung aus ausführen. Weitere Informationen zu den Shell-Befehlen und deren Aufruf erhalten Sie, wenn Sie an der HBase Eingabeaufforderung `help` eingeben und die Eingabetaste drücken.

Erstellen einer Tabelle

Mit dem folgenden Befehl wird eine Tabelle namens „t1“ erstellt, die eine Spaltengruppe mit dem Namen „f1“ enthält.

```
hbase(main):001:0>create 't1', 'f1'
```

Eingeben eines Werts

Mit dem folgenden Befehl wird der Wert „v1“ für die Zeile „r1“ in Tabelle „t1“ und Spalte „f1“ eingetragen.

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

Abrufen eines Werts

Mit dem folgenden Befehl werden die Werte für Zeile „r1“ in Tabelle „t1“ abgerufen.

```
hbase(main):001:0>get 't1', 'r1'
```

Löschen einer Tabelle

Der folgende Befehl löscht und löscht die Tabelle „t1“.

```
hbase(main):001:0>drop 'ns1:t1', false
```

Der boolesche Wert gibt an, ob Sie Ihre Tabelle archivieren möchten oder nicht. Sie können ihn also auf `true` festlegen, wenn Sie sie speichern möchten. Sie können `drop 'ns1:t1'` auch ohne booleschen Wert ausführen, um die Tabelle zu archivieren.

Greifen Sie mit Hive auf HBase Tabellen zu

HBase und [Apache Hive](#) sind eng integriert, sodass Sie massive Parallelverarbeitungs-Workloads direkt für die in gespeicherten Daten ausführen können. HBase Um Hive mit zu verwenden HBase, können Sie sie normalerweise auf demselben Cluster starten. Sie können Hive jedoch auch HBase auf separaten Clustern starten. Wenn Sie Hive HBase und Hive getrennt auf verschiedenen Clustern ausführen, kann dies die Leistung verbessern, da so jede Anwendung Clusterressourcen effizienter nutzen kann.

Die folgenden Verfahren zeigen, wie Sie mithilfe von Hive eine Verbindung HBase zu einem Cluster herstellen.

Note

Sie können einen Hive-Cluster nur mit einem einzelnen HBase Cluster verbinden.

Um Hive zu verbinden HBase

1. Erstellen Sie separate Cluster mit Hive und HBase installiertem Hive oder erstellen Sie einen einzelnen Cluster, auf dem HBase sowohl Hive als auch Hive installiert sind.

2. Wenn Sie separate Cluster verwenden, ändern Sie Ihre Sicherheitsgruppen so, dass die HBase Hive-Ports zwischen diesen beiden Primärknoten geöffnet sind.
3. Wird verwendet SSH, um eine Verbindung zum primären Knoten für den Cluster herzustellen, auf dem Hive installiert ist. Weitere Informationen finden Sie unter [Connect mit dem primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.
4. Starten Sie die Hive-Shell mit dem folgenden Befehl.

```
hive
```

5. (Optional) Sie müssen dies nicht tun, wenn sich Hive HBase und Hive auf demselben Cluster befinden. Connect den HBase Client auf Ihrem Hive-Cluster mit dem HBase Cluster, der Ihre Daten enthält. Im folgenden Beispiel *public-DNS-name* wird durch den öffentlichen DNS Namen des primären Knotens des HBase Clusters ersetzt, zum Beispiel: `ec2-50-19-76-67.compute-1.amazonaws.com`.

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. Fahren Sie mit der Ausführung von Hive-Abfragen für Ihre HBase Daten wie gewünscht fort oder sehen Sie sich das nächste Verfahren an.

Um auf HBase Daten von Hive zuzugreifen

- Nachdem die Verbindung zwischen Hive und HBase Clustern hergestellt wurde (wie im vorherigen Verfahren gezeigt), können Sie auf die im HBase Cluster gespeicherten Daten zugreifen, indem Sie eine externe Tabelle in Hive erstellen.

Wenn das folgende Beispiel über die Hive-Eingabeaufforderung auf dem Primärknoten ausgeführt wird, wird eine externe Tabelle erstellt, die auf Daten verweist, die in einer HBase Tabelle mit dem Namen gespeichert sind. `inputTable` Anschließend können Sie `inputTable` in Hive-Anweisungen referenzieren, um im Cluster gespeicherte Daten abzufragen und zu ändern. HBase

```
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 ;
```

Einen fortgeschritteneren Anwendungsfall und ein Beispiel für die Kombination HBase mit Hive finden Sie im AWS Big-Data-Blogbeitrag [Combine No SQL and Massively Parallel Analytics using Apache HBase and Apache Hive on Amazon](#). EMR

Schnappschüsse verwenden HBase

HBase verwendet eine integrierte [Snapshot-Funktionalität](#), um einfache Backups von Tabellen zu erstellen. In EMR Clustern können diese Backups mit nach Amazon S3 exportiert werden. Sie können mithilfe der HBase Shell einen Snapshot auf dem primären Knoten erstellen. In diesem Thema erfahren Sie, wie Sie diese Befehle interaktiv mit der Shell oder in einem Schritt ausführen, indem Sie entweder das AWS CLI oder AWS SDK for Java verwenden `command-runner.jar`. Weitere Informationen zu anderen Arten von HBase Backups finden Sie in der HBase Dokumentation unter [HBaseBackup](#).

Erstellen eines Snapshots mit einer Tabelle

```
hbase snapshot create -n snapshotName -t tableName
```

Verwenden von `command-runner.jar` aus dem 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

Wenn Ihr Snapshot-Name nicht eindeutig ist, schlägt der Erstellvorgang mit dem Rückgabecode -1 oder 255 fehl. Gegebenenfalls wird aber keine Fehlermeldung mit den Details zum Fehler angezeigt. Um den gleichen Snapshot-Namen zu verwenden, löschen Sie ihn und erstellen Sie ihn dann erneut.

Löschen eines Snapshots

```
hbase shell
>> delete_snapshot 'snapshotName'
```

Anzeigen von Snapshot-Informationen

```
hbase snapshot info -snapshot snapshotName
```

Einen Snapshot nach Amazon S3 exportieren

⚠ Important

Wenn Sie beim Exportieren eines Snapshots keinen `-mappers` Wert angeben, HBase verwendet eine willkürliche Berechnung, um die Anzahl der Mapper zu ermitteln. Dieser Wert kann je nach Tabellengröße sehr hoch sein, was sich negativ auf beim Export laufende Aufträge auswirkt. Aus diesem Grund empfehlen wir, den Parameter `-mappers`, den Parameter `-bandwidth` (der den Bandbreitenverbrauch in Megabyte pro Sekunde angibt) oder beide anzugeben, um die von der Exportoperation verwendeten Cluster-Ressourcen zu begrenzen. Alternativ können Sie die Operation zum Exportieren des Snapshots während eines Zeitraums mit geringer Auslastung ausführen.

```
hbase snapshot export -snapshot snapshotName \  
-copy-to s3://bucketName/folder -mappers 2
```

Unter Verwendung `command-runner.jar` von: AWS CLI

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \  
--command-runner-jar-path s3://bucketName/command-runner.jar --command "hbase snapshot export -snapshot snapshotName -copy-to s3://bucketName/folder -mappers 2"
```

```
--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");
```

Importieren eines Snapshots von Amazon S3

Obwohl es sich um einen Import handelt, ist die hier verwendete HBase Option immer noch gültig export.

```
sudo -u hbase hbase snapshot export \
-D hbase.rootdir=s3://bucketName/folder \
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

Verwendung command-runner.jar von 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");
```

Eine Tabelle aus Snapshots in der HBase Shell wiederherstellen

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBaseunterstützt derzeit nicht alle Snapshot-Befehle in der HBase Shell. Beispielsweise gibt es keine HBase Befehlszeilenoption zum Wiederherstellen eines Snapshots, sodass Sie ihn in einer Shell wiederherstellen müssen. Dies bedeutet, dass `command-runner.jar` einen Bash-Befehl ausführen muss.

Note

Da der hier verwendete Befehl `lautetecho`, ist es möglich, dass Ihr Shell-Befehl auch dann fehlschlägt, wenn der von Amazon ausgeführte Befehl einen `0` Exit-Code EMR zurückgibt. Überprüfen Sie die Schrittprotokolle, wenn Sie einen Shell-Befehl als Schritt ausführen möchten.

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

Nachfolgend finden Sie den Schritt mit der AWS CLI. Erstellen Sie zunächst die folgende `snapshot.json`-Datei:

```
[
  {
    "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");
```

Konfigurieren HBase

Obwohl die HBase Standardeinstellungen für die meisten Anwendungen funktionieren sollten, können Sie Ihre HBase Konfigurationseinstellungen ändern. Verwenden Sie dazu die Eigenschaften von HBase Konfigurationsklassifizierungen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Im folgenden Beispiel wird ein Cluster mit einem alternativen HBase Stammverzeichnis erstellt `myConfig.json`, das auf einer in Amazon S3 gespeicherten Konfigurationsdatei basiert.

Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Die Datei `myConfig.json` gibt die `hbase.rootdir`-Eigenschaft für die `hbase-site`-Konfigurationsklassifizierung an, wie im folgenden Beispiel gezeigt. Ersetzen `ip-XXX-XX-XX-XXX.ec2.internal` mit dem internen DNS Hostnamen des primären Knotens des Clusters.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {
```

```
"hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/  
myCustomHBaseDir"  
  }  
}  
]
```

Note

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Dazu verwenden Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Änderungen an der Speicherzuweisung in YARN

HBase wird nicht als YARN Anwendung ausgeführt. Daher ist es notwendig, den zugewiesenen Speicher YARN und die zugehörigen Anwendungen neu zu berechnen, was zu einer Verringerung des Gesamtspeichers führt, der bei der Installation zur YARN Verfügung steht. HBase Sie sollten dies berücksichtigen, wenn Sie planen, YARN Anwendungen gemeinsam HBase auf denselben Clustern zu platzieren. Bei den Instance-Typen mit weniger als 64 GB Arbeitsspeicher steht die Hälfte des Speichers zur Verfügung NodeManager, der dann dem HBase RegionServer zugewiesen wird. Bei Instance-Typen mit einem Arbeitsspeicher von mehr als 64 GB ist der HBase RegionServer Arbeitsspeicher auf 32 GB begrenzt. Als allgemeine Regel gilt, dass das YARN Einstellen des Speichers ein Vielfaches des MapReduce Reducer-Task-Speichers ist.

Die Tabellen in der Tabelle [Standardwerte für die Aufgabenkonfigurationseinstellungen](#) zeigen die Änderungen an den YARN Einstellungen auf der Grundlage des Speicherbedarfs für HBase.

HBase Portnummern

Einige der ausgewählten Portnummern HBase unterscheiden sich von den Standardnummern. Im Folgenden finden Sie Schnittstellen und Ports für HBase Amazon EMR.

HBase-Anschlüsse

Schnittstelle	Port	Protokoll
HMaster	16000	TCP
HMasterUI	16010	HTTP
RegionServer	16020	TCP
RegionServer Informationen	16030	HTTP
RETServert	8070	HTTP
RETServertBenutzeroberfläche	8085	HTTP
Thrift-Server	9090	TCP
Thrift-Server-UI	9095	HTTP

Wichtig

Das `kms-http-port` ist 9700 und das `kms-admin-port` ist 9701 in der EMR Amazon-Release-Version 4.6.0 und höher.

HBase-Seiteneinstellungen zur Optimierung

Sie können einzelne oder alle HBase Site-Einstellungen festlegen, um den HBase Cluster für die Arbeitslast Ihrer Anwendung zu optimieren. Wir empfehlen die folgenden Einstellungen als Ausgangspunkt.

`zookeeper.session.timeout`

Der Timeout-Standardwert beträgt 40 Sekunden (40 000 ms). Wenn ein Regionsserver ausfällt, gibt dieser Wert an, wie lange es dauert, bis der Master-Server die Abwesenheit des Regionsservers erkennt und mit der Wiederherstellung beginnt. Um die Wiederherstellung des Master-Servers zu beschleunigen, können Sie diesen Wert auf einen kürzeren Zeitraum einstellen. Im folgenden Beispiel werden 30 Sekunden oder 30 000 ms verwendet:


```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "zookeeper.session.timeout": "30000"
    }
  }
]
```

hbase.regionserver.handler.count

Dadurch wird die Anzahl der Threads definiert, die der Regionsserver offen hält, um Tabellenanforderungen zu verarbeiten. Die Standardeinstellung 10 ist niedrig, damit Benutzer ihre Regionsserver nicht überlasten, wenn sie zu große Schreibpuffer mit vielen gleichzeitigen Clients verwenden. Als Faustregel gilt, diese Zahl niedrig zu halten, wenn sich die Nutzlast pro Anfrage dem MB-Bereich nähert (große Puts, Scans mit großem Cache) und hoch, wenn die Payload klein ist (Gets, Small Puts/ICVs, Deletes). Im folgenden Beispiel wird die Anzahl der offenen Threads auf 30 erhöht:

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.regionserver.handler.count": "30"
    }
  }
]
```

hbase.hregion.max.filesize

Dieser Parameter steuert die Größe (in Byte) der einzelnen Regionen. Standardmäßig ist der Wert eingestellt 1073741824. Wenn Sie viele Daten in Ihren HBase Cluster schreiben und dies zu häufigem Splitting führt, können Sie diese Größe erhöhen, um einzelne Regionen zu vergrößern. Dadurch werden zwar Aufteilungen reduziert, es braucht jedoch mehr Zeit, um die Workloads in den Regionen zwischen den Servern zu verteilen.

```
[
  {
    "Classification": "hbase-site",
```

```
"Properties": {
  "hbase.hregion.max.filesize": "1073741824"
}
]
```

hbase.hregion.memstore.flush.size

Dieser Parameter steuert die maximale Größe des Memstore (in Byte), bevor er auf den Datenträger ausgelagert wird. Standardmäßig ist dieser 134217728. Wenn Ihr Workload aus Schreibvorgängen in kurzen Spitzen besteht, sollten Sie dieses Limit erhöhen, damit alle Schreibvorgänge während der Spitzenauslastung im Arbeitsspeicher verbleiben und zu einem späteren Zeitpunkt auf den Datenträger ausgelagert werden. Dies kann die Leistung bei Auslastungsspitzen verbessern.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.memstore.flush.size": "134217728"
    }
  }
]
```

Sehen Sie sich die HBase Benutzeroberfläche an

Note

Die HBase Benutzeroberfläche verwendet standardmäßig unsichere HTTP Verbindungen. Um secure HTTP (HTTPS) zu aktivieren, setzen Sie `true` in Ihrer [HBaseKonfiguration](#) die `hbase.ssl.enabled` Eigenschaft für die `hbase-site` Klassifizierung auf. Weitere Informationen zur Verwendung von secure HTTP (HTTPS) für die HBase Weboberfläche finden Sie im [Apache HBase Reference Guide](#).

HBase bietet eine webbasierte Benutzeroberfläche, mit der Sie Ihren HBase Cluster überwachen können. Wenn Sie es HBase auf Amazon ausführen EMR, läuft das Webinterface auf dem primären Knoten und kann mithilfe von Portweiterleitung, auch bekannt als „Erstellen eines SSH Tunnels“, aufgerufen werden.

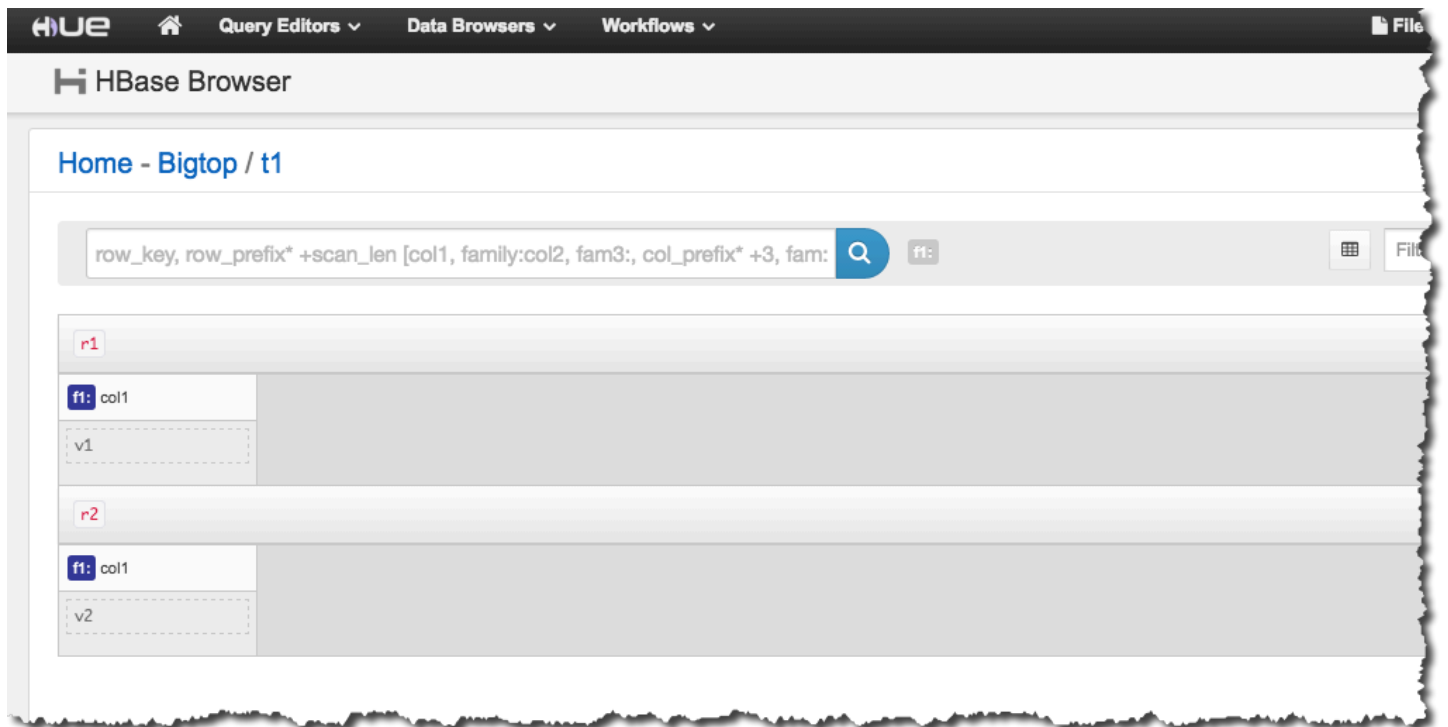
Um die HBase Benutzeroberfläche anzuzeigen

1. Wird verwendet SSH, um einen Tunnel in den Primärknoten zu tunneln und eine sichere Verbindung herzustellen. Weitere Informationen finden Sie unter [Option 2, Teil 1: Einen SSH Tunnel zum Primärknoten mithilfe dynamischer Portweiterleitung einrichten](#) im Amazon EMR Management Guide.
2. Installieren Sie einen Webbrowser mit einem Proxy-Tool, z. B. dem FoxyProxy Plug-in für Firefox, um einen SOCKS Proxy für AWS Domains zu erstellen. Weitere Informationen finden Sie unter [Option 2, Teil 2: Proxyeinstellungen konfigurieren, um Websites anzuzeigen, die auf dem primären Knoten gehostet werden](#), im Amazon EMR Management Guide.
3. Wenn der Proxy eingerichtet und die SSH Verbindung geöffnet ist, können Sie die HBase Benutzeroberfläche aufrufen, indem Sie ein Browserfenster mit `http://öffnenmaster-public-dns-name:16010/master-status`, wo ~~master-public-dns-name~~ ist die öffentliche DNS Adresse des primären Knotens des Clusters.

The screenshot shows the Apache HBase web interface. At the top, there is a navigation bar with the following links: Home, Table Details, Local Logs, Log Level, Debug Dump, Metrics Dump, and HBase Configuration. Below the navigation bar, the page title is "Master [redacted].ec2.internal". Underneath, there is a section titled "Region Servers" with a sub-section "Base Stats". Below this, there is a table with the following columns: ServerName, Start time, Version, Requests Per Second, and Num. P. The table contains two rows of data and a total row.

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

Sie können es auch HBase in Hue ansehen. Das folgende Beispiel zeigt die Tabelle t1, erstellt in [Verwenden der HBase Shell](#):



Weitere Informationen zu Hue finden Sie unter [Hue](#).

Anzeige von HBase-Protokolldateien

HBase schreibt im Rahmen dieses Vorgangs Protokolldateien mit Details zu Konfigurationseinstellungen, Daemon-Aktionen und Ausnahmen. Diese Protokolldateien können sowohl für das Debuggen von Problemen HBase als auch für die Leistungsverfolgung nützlich sein.

Wenn Sie Ihren Cluster so konfigurieren, dass die Protokolldateien dauerhaft in Amazon S3 gespeichert werden, sollten Sie wissen, dass Protokolle alle fünf Minuten in Amazon S3 geschrieben werden. Deshalb kann es zu einer kleinen Verzögerung kommen, bis die neuesten Protokolldateien verfügbar sind.

Um HBase Protokolle auf dem primären Knoten anzuzeigen

- Sie können die aktuellen HBase Protokolle anzeigenSSH, indem Sie eine Verbindung zum primären Knoten herstellen und zum `/var/log/hbase` Verzeichnis navigieren. Diese Protokolle sind nicht verfügbar, nachdem der Cluster beendet wurde, es sei denn, Sie aktivieren beim Starten des Clusters die Protokollierung nach Amazon S3.

So zeigen Sie HBase Protokolle auf Amazon S3 an

- Um auf HBase Protokolle und andere Cluster-Protokolle in Amazon S3 zuzugreifen und diese nach der Clusterbeendigung verfügbar zu haben, geben Sie einen Amazon S3 S3-Bucket an, der diese Protokolle empfängt, wenn Sie den Cluster erstellen. Dies erfolgt mithilfe der Option `--log-uri`. Weitere Informationen zur Aktivierung der Protokollierung für Ihren Cluster finden Sie unter [Protokollierung und Debugging konfigurieren \(optional\)](#) im Amazon EMR Management Guide.

Monitor HBase mit Ganglia

Das Open-Source-Projekt Ganglia ist ein skalierbares, verteiltes System zur Überwachung von Clustern und Grids, das zugleich die Auswirkungen auf die Leistung minimiert. Wenn Sie Ganglia in Ihrem Cluster aktivieren, können Sie Berichte erstellen und die Leistung des Clusters als Ganzes betrachten. Ebenso können Sie die Leistung einzelner Knoten-Instances überprüfen. Weitere Informationen zum Open-Source-Projekt Ganglia finden Sie unter <http://ganglia.info/>. Weitere Informationen zur Verwendung von Ganglia mit EMR Amazon-Clustern finden Sie unter [Ganglia](#).

Nachdem der Cluster mit konfigurierter Ganglia gestartet wurde, können Sie über die grafische Oberfläche, die auf dem primären Knoten ausgeführt wird, auf die Ganglia-Diagramme und -Berichte zugreifen.

Ganglia speichert Protokolldateien auf dem primären Knoten im Verzeichnis `/mnt/var/lib/ganglia/rrds/`. Frühere Versionen von Amazon EMR können Protokolldateien im `/var/log/ganglia/rrds/` Verzeichnis speichern.

Um einen Cluster für Ganglia zu konfigurieren und HBase den AWS CLI

- Verwenden Sie einen `create-cluster`-Befehl ähnlich dem folgenden:

```
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

Wenn die standardmäßige EMR Amazon-Servicerolle und das EC2 Amazon-Instanzprofil nicht existieren, tritt ein Fehler auf. Verwenden Sie den Befehl `aws emr create-default-roles`, um die Rolle und das Profil zu erstellen, und versuchen Sie es erneut.

Weitere Informationen finden Sie unter [EMRAmazon-Befehle in der AWS CLI](#).

Um HBase Metriken in der Ganglia-Weboberfläche anzuzeigen

1. Wird verwendetSSH, um einen Tunnel in den Primärknoten zu tunneln und eine sichere Verbindung herzustellen. Weitere Informationen finden Sie unter [Option 2, Teil 1: Einen SSH Tunnel zum Primärknoten mithilfe dynamischer Portweiterleitung einrichten](#) im Amazon EMR Management Guide.
2. Installieren Sie einen Webbrowser mit einem Proxy-Tool, z. B. dem FoxyProxy Plug-in für Firefox, um einen SOCKS Proxy für AWS Domains zu erstellen. Weitere Informationen finden Sie unter [Option 2, Teil 2: Proxyeinstellungen konfigurieren, um Websites anzuzeigen, die auf dem primären Knoten gehostet](#) werden, im Amazon EMR Management Guide.
3. Wenn der Proxy eingerichtet und die SSH Verbindung geöffnet ist, können Sie die Ganglia-Metriken anzeigen, indem Sie ein Browserfenster mit `http://öffnenmaster-public-dns-name/Ganglia/` öffnen, wo `master-public-dns-name` ist die öffentliche DNS Adresse des Masterservers im Cluster. HBase

Um Ganglia-Protokolldateien auf dem primären Knoten anzuzeigen

- Wenn der Cluster noch läuft, können Sie auf die Protokolldateien zugreifen, indem Sie eine Verbindung SSH zum primären Knoten herstellen und zum `/mnt/var/lib/ganglia/rrds/` Verzeichnis navigieren. Für EMR 3.x navigieren Sie zum `/var/log/ganglia/rrds` Verzeichnis. Weitere Informationen finden Sie unter [Connect mit dem primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.

So zeigen Sie Ganglia-Protokolldateien in Amazon S3 an

- Die Ganglia-Protokolldateien werden nicht automatisch nach Amazon S3 geschrieben, auch wenn Sie die Protokollierung für den Cluster aktivieren. Um die Ganglia-Protokolldateien in Amazon S3 anzuzeigen, müssen Sie die Protokolle manuell per Push von `/mnt/var/lib/ganglia/rrds/` in den S3-Bucket übertragen.

Migration von früheren Versionen HBase

Informationen zur Migration von Daten aus einer früheren HBase Version finden Sie unter [Upgrade, HBase Versionsnummer und Kompatibilität](#) im Apache HBase Reference Guide. Möglicherweise müssen Sie den Anforderungen für ein Upgrade von Versionen vor 1.0 besondere Aufmerksamkeit schenken. HBase

HBase Versionsverlauf

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon HBase enthalten sindEMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

HBase Versionsinformationen

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-7.2.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-ma-pred, hadoop-yarn-nodema-nager, hadoop-yarn-resourcemanager, hadoop-yarn-

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
		<p>timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server</p>
emr-5.36,2	1.4.13	<p>emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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</p>

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-6.15.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-6.11.1	2.4,15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-6.10.0	2.4,15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma-pred, hadoop-yarn-nodema-nager, hadoop-yarn-resour-cemanager, hadoop-yarn-timeline-server, hbase-hma-ster, hbase-client, hbase-reg-ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope-rator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.6.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-6.4.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.3.1	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-6.2.0	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.1.1	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-6.0.0	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.34.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.33.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.32.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.32.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.30.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.25.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.24.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.23.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.22.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.2	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.21.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.17.2	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.1	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.17.0	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.16.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.16.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.15.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.15.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.2	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.12.3	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.2	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.11.4	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.10.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.9.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.9.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.8.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.7.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.7.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.6.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.6.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.4	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.5.3	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.2	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.3.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.3	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-5.0.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.6	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-regionserver, 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-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-4.9.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit 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

Amazon EMR Release-Etikett	HBaseAusführung	Komponenten wurden installiert mit HBase
emr-4.6.1	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.6.0	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Apache HCatalog

HCatalog ist ein Tool, mit dem Sie in Pig-, Spark SQL - und/oder benutzerdefinierten Anwendungen auf Hive-Metastore-Tabellen zugreifen können. MapReduce HCatalog verfügt über eine REST Benutzeroberfläche und einen Befehlszeilenclient, mit dem Sie Tabellen erstellen oder andere Operationen ausführen können. Anschließend schreiben Sie Ihre Anwendungen, um mithilfe von HCatalog Bibliotheken auf die Tabellen zuzugreifen. Weitere Informationen finden Sie unter [Verwendung von HCatalog](#). HCatalog ist in der EMR Amazon-Release-Version 4.4.0 und höher enthalten.

HCatalog auf Amazon EMR Release Version 5.8.0 und höher unterstützt die Verwendung von AWS Glue Data Catalog als Metastore für Hive. Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive](#).

In der folgenden Tabelle sind die Version von HCatalog aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. HCatalog

Informationen zur Version der Komponenten, mit denen HCatalog in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 6.15.0](#).

HCatalog Versionsinformationen für emr-6.15.0

EMR Amazon-Freigabeetikett	HCatalog Ausführung	Mit HCatalog installierte Komponenten
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-s

EMRAmazon-Freigabeetikett	HCatalogAusführung	Mit HCatalog installierte Komponenten
		erver, hive-client, mariadb-server

In der folgenden Tabelle sind die Version von HCatalog aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. HCatalog

Informationen zur Version der Komponenten, mit denen HCatalog in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

HCatalogVersionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	HCatalogAusführung	Mit HCatalog installierte Komponenten
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

Themen

- [Einen Cluster erstellen mit HCatalog](#)
- [Verwenden HCatalog](#)
- [Beispiel: Erstellen Sie eine HCatalog Tabelle und schreiben Sie mit Pig in sie](#)

- [HCatalog Versionsverlauf](#)

Einen Cluster erstellen mit HCatalog

Obwohl HCatalog es im Hive-Projekt enthalten ist, müssen Sie es als eigene Anwendung installieren.

Um einen Cluster zu starten, der über die Konsole HCatalog installiert ist

Das folgende Verfahren erstellt einen Cluster mit HCatalog installiertem. Weitere Informationen zum Erstellen von Clustern mithilfe der Konsole, einschließlich erweiterter Optionen, finden Sie unter [Planen und Konfigurieren von Clustern](#) im Amazon EMR Management Guide.

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create cluster (Cluster erstellen) aus, um Quick Create (Schnellerstellung) zu verwenden.
3. Wählen Sie im Feld Software Configuration (Softwarekonfiguration) die Option Amazon Release Version emr-4.4.0 oder höher aus.
4. Wählen Sie im Feld „Anwendungen auswählen“ entweder „Alle Anwendungen“ oder aus HCatalog.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Um einen Cluster HCatalog mit dem zu starten AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl:

Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
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
```

Verwenden HCatalog

Sie können ihn HCatalog in verschiedenen Anwendungen verwenden, die den Hive-Metastore verwenden. Die Beispiele in diesem Abschnitt zeigen, wie Sie eine Tabelle erstellen und sie im Kontext von Pig und Spark verwenden. SQL

Deaktivieren Sie direktes Schreiben bei der Verwendung von HCatalog HStorer

Wenn eine Anwendung in [HCatStorer](#) eine in Amazon S3 gespeicherte HCatalog Tabelle schreibt, deaktivieren Sie die Direktschreibfunktion von AmazonEMR. Deaktivieren Sie beispielsweise Direct Write, wenn Sie den STORE Pig-Befehl verwenden oder wenn Sie Sqoop-Jobs ausführen, die HCatalog Tabellen in Amazon S3 schreiben. Sie können die Direct Write-Funktion deaktivieren, indem Sie die Konfigurationen `mapred.output.direct.NativeS3FileSystem` und `mapred.output.direct.EmrFileSystem` auf `false` setzen. Das folgende Beispiel zeigt, die diese Konfigurationen mit Java festgelegt werden.

```
Configuration conf = new Configuration();
conf.set("mapred.output.direct.NativeS3FileSystem", "false");
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

Erstellen Sie eine Tabelle mit den HCat CLI und verwenden Sie diese Daten in Pig

Erstellen Sie das folgende Skript, `impressions.q`, auf Ihrem 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');
```

Führen Sie das Skript aus mit dem 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
```

Öffnen Sie die Grunt-Shell und greifen Sie auf die Daten in `impressions` zu:

```
% 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,wcVWWTascoPbGt6bdqDbuWTPPHg0Ps,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,q2Awfnpe0JAvhInaIp0VGx9Kts0oPO,s3HvTf1PB8JIE0IuM6h0EebWp0tJV,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,c362vpAB0soPKGHR543cj6TRwNe0Gn,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)
```

Zugreifen auf die Tabelle mit Spark SQL

In diesem Beispiel wird `DataFrame` aus der Tabelle, die im ersten Beispiel erstellt wurde, ein Spark-Objekt erstellt und die ersten 20 Zeilen angezeigt:

```
% 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|m9nwdo67Nx6q2kI25...|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|cjbTpruoaiEtqLuMX...|XwlohBSs8Ipxs1bRa...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610804000|Ms3eJHNAEItpxvimd...|4SIj4pGmgVL1625BD...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610872000|h5bccHX6wJReDi1jL...|EFAWiiBdVfnxwAMWP...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610365000|874NBpGmxNFfxEPKM...|xSvE4XtGbdTXPF2Lb...|cartoonnetwork.com|
Mozilla/5.0 (Maci...|eWDEVVUphlnRa273j...| 22.91.173.232|2009-04-13-08-05|
|  1239610348000|X8gISpUTSsqh1A5reS...|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|Wf1svKIg0qfIE5KwR...|TJHd1VBspNcua0XPn...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
| 1239610768000|4MJR0XxiVCU1ueXKV...|10hGWmbvKf8ajoU8a...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
| 1239610832000|gWIrpDiN57i3sHatv...|RNL4C7xPi3tdar2Uc...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
| 1239610789000|pTne9k62kJ14QViXI...|RVxJVIQousjxUVI3r...| pixnet.net|
Mozilla/5.0 (Maci...|1bG0KiBD2xmui90kF...| 33.176.101.80|2009-04-13-08-05|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

```

```
scala>
```

Beispiel: Erstellen Sie eine HCatalog Tabelle und schreiben Sie mit Pig in sie

Sie können eine HCatalog Tabelle erstellen und HCatStorer mit Apache Pig in sie schreiben, indem Sie eine Datenquelle in Amazon S3 verwenden. HCatalogerfordert, dass Sie Direct Write deaktivieren, andernfalls schlägt der Vorgang unbemerkt fehl. Legen Sie sowohl die Konfiguration `mapred.output.direct.NativeS3FileSystem` als auch `mapred.output.direct.EmrFileSystem` entweder mit der `false`-Klassifizierung oder manuell innerhalb der Grunt-Shell auf `mapred-site` fest. Das folgende Beispiel zeigt eine Tabelle HCatCLI, die mit den Befehlen erstellt wurde, gefolgt von Befehlen, die in der Grunt-Shell ausgeführt werden, um die Tabelle aus einer Beispieldatendatei in Amazon S3 zu füllen.

Um dieses Beispiel auszuführen, [stellen Sie mithilfe von eine Verbindung zum Master-Knoten](#) her. SSH

Erstellen Sie eine HCatalog Skriptdatei mit dem folgenden Inhalt, die eine HCatalog Tabelle mit dem Namen `ersteltwikicount.wikicount.q`

```
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';
```

Verwenden Sie einen HCat CLI Befehl, um das Skript aus der Datei auszuführen.

```
hcat -f wikicount.q
```

Im nächsten Schritt starten Sie die Grund-Shell mit der Option `-useHCatalog`, legen Konfigurationen zum Deaktivieren der Direct Write-Funktion fest, laden Daten von einem S3-Speicherort und schreiben die Ergebnisse in die Wikicount-Tabelle.

```
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();
```

HCatalogVersionsverlauf

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon HCatalog enthalten sindEMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

HCatalogVersionsinformationen

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client,

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
		<p>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</p>
emr-5.36,2	2.3.9	<p>emrfs, emr-ddb, emr-goodies, 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</p>

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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
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

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Amazon EMR Release-Etikett	HCatalogAusführung	Komponenten wurden installiert mit HCatalog
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Apache Hive

Bei Hive handelt es sich um ein Open-Source-Data-Warehouse und Analyse-Paket von Datensätzen, das auf einem Hadoop-Cluster ausgeführt wird. Hive-Skripte verwenden eine SQL ähnliche Sprache namens Hive QL (Abfragesprache), die Programmiermodelle abstrahiert und typische Data Warehouse-Interaktionen unterstützt. Mit Hive können Sie die Komplexität vermeiden, die mit dem Schreiben von Tez-Jobs verbunden ist, die auf gerichteten azyklischen Graphen (DAGs) oder MapReduce Programmen in einer Computersprache niedrigerer Stufe wie Java basieren.

Hive erweitert das Paradigma um Serialisierungsformate. SQL Sie können die Abfrageverarbeitung auch anpassen, indem Sie ein Tabellenschema erstellen, das mit Ihren Daten übereinstimmt, ohne die eigentlichen Daten zu bearbeiten. Hive-Tabellenwerte unterstützen zwar SQL nur primitive Werttypen (wie Datumsangaben, Zahlen und Zeichenketten), sind jedoch strukturierte Elemente wie JSON Objekte, beliebige benutzerdefinierte Datentypen oder in Java geschriebene Funktionen.

Weitere Informationen zu Hive finden Sie unter <http://hive.apache.org/>.

In der folgenden Tabelle sind die Version von Hive aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Hive EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hive installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Hive-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Hive-Version	Mit Hive installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-

EMRAmazon-Freigabeetikett	Hive-Version	Mit Hive installierte Komponenten
		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

In der folgenden Tabelle sind die Version von Hive aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Hive EMR installiert.

Die Version der Komponenten, die mit Hive in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Hive-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Hive-Version	Mit Hive installierte Komponenten
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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-

EMRAmazon-Freigabeetikett	Hive-Version	Mit Hive installierte Komponenten
		worker, zookeeper-client, zookeeper-server

In der folgenden Tabelle sind die Version von Hive aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Hive EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hive installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Hive-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Hive-Version	Mit Hive installierte Komponenten
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

Ab Amazon EMR 5.18.0 können Sie das Amazon EMR Artifact Repository verwenden, um Ihren Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die in bestimmten Amazon-Versionen verfügbar sind. EMR Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Themen

- [Unterschiede und Überlegungen zu Hive bei Amazon EMR](#)
- [Konfigurieren eines externen Metastores für Hive](#)
- [Den Hive-JDBC-Treiber verwenden](#)
- [Verbessern der Hive-Leistung](#)
- [Verwendung von Hive Live Long und Process \(\) LLAP](#)
- [Verschlüsselung im Hive](#)
- [Versionsverlauf von Hive](#)

Unterschiede und Überlegungen zu Hive bei Amazon EMR

Unterschiede zwischen Apache Hive auf Amazon EMR und Apache Hive

[In diesem Abschnitt werden die Unterschiede zwischen Hive on Amazon EMR und den Standardversionen von Hive beschrieben, die unter <http://svn.apache.org/viewvc/hive/branches/> verfügbar sind.](#)

Hive-Autorisierung

Amazon EMR unterstützt die [Hive-Autorisierung](#) fürHDFS, aber nicht für EMRFS Amazon S3. EMRAmazon-Cluster werden standardmäßig mit deaktivierter Autorisierung ausgeführt.

Dateizusammenführung mit Hive in Amazon S3

Apache Hive führt kleine Dateien am Ende eines Map-Auftrags zusammen, wenn `hive.merge.mapfiles` auf "true" gesetzt ist und die Zusammenführung nur dann ausgelöst wird, wenn die durchschnittliche Ausgabegröße des Auftrags kleiner als die Einstellung `hive.merge.smallfiles.avgsize` ist. Amazon EMR Hive hat genau das gleiche Verhalten, wenn der endgültige Ausgabepfad angegeben ist. HDFS Wenn sich der Ausgabepfad in Amazon S3 befindet, wird der Parameter `hive.merge.smallfiles.avgsize` jedoch ignoriert. In diesem Fall wird die Zusammenführungsaufgabe immer dann ausgelöst, wenn `hive.merge.mapfiles` auf `true` gesetzt ist.

ACIDTransaktionen und Amazon S3

Amazon EMR 6.1.0 und höher unterstützt Hive-Transaktionen ACID (Atomicity, Consistency, Isolation, Durability), sodass sie den ACID Eigenschaften einer Datenbank entsprechen. Mit dieser

Funktion können Sie, INSERT UPDATEDELETE, und MERGE Operationen in von Hive verwalteten Tabellen mit Daten in Amazon Simple Storage Service (Amazon S3) ausführen.

Hive Live Long und Process () LLAP

[LLAPFunktionen](#), die in Version 2.0 des Standard-Apache Hive hinzugefügt wurden, werden in Hive 2.1.0 auf EMR Amazon-Version 5.0 nicht unterstützt.

EMRAmazon-Version 6.0.0 und höher unterstützt die Funktionen Live Long und Process (LLAP) für Hive. Weitere Informationen finden Sie unter Hive [verwenden](#). LLAP

Unterschiede in Hive zwischen EMR Amazon-Release-Version 4.x und 5.x

In diesem Abschnitt werden die Unterschiede behandelt, die Sie berücksichtigen sollten, bevor Sie eine Hive-Implementierung von Hive-Version 1.0.0 auf Amazon-Version 4.x auf Hive 2.x auf EMR Amazon-Version 5.x migrieren. EMR

Betriebliche Unterschiede und Überlegungen

- Support für [Transaktionen ACID \(Atomizität, Konsistenz, Isolierung und Haltbarkeit\)](#) hinzugefügt: Dieser Unterschied zwischen Hive 1.0.0 auf Amazon EMR 4.x und dem Standard-Apache Hive wurde beseitigt.
- Direkte Schreibvorgänge auf Amazon S3 wurden beseitigt: Dieser Unterschied zwischen Hive 1.0.0 auf Amazon EMR und dem Standard-Apache Hive wurde beseitigt. Hive 2.1.0 auf Amazon EMR Version 5.x erstellt, liest aus und schreibt in temporäre Dateien, die in Amazon S3 gespeichert sind. Um aus derselben Tabelle zu lesen und in dieselbe Tabelle zu schreiben, müssen Sie daher als Workaround keine temporäre Tabelle mehr im lokalen HDFS Dateisystem des Clusters erstellen. Wenn Sie versionierte Buckets verwenden, stellen Sie sicher, dass diese temporären Dateien wie unten beschrieben verwaltet werden.
- Verwalten Sie temporäre Dateien, wenn Sie versionierte Amazon-S3-Buckets verwenden: Wenn Sie Hive-Abfragen ausführen, bei denen das Ziel der generierten Daten Amazon S3 ist, werden viele temporäre Dateien und Verzeichnisse erstellt. Hierbei handelt es sich, wie bereits beschrieben, um ein neues Verhalten. Wenn Sie versionierte S3-Buckets verwenden, wird Amazon S3 von diesen temporären Dateien überfrachtet und es fallen Kosten an, wenn sie nicht gelöscht werden. Passen Sie Ihre Lebenszyklusregeln an, damit Daten mit einem /_tmp-Präfix nach einem kurzen Zeitraum, wie z. B. nach fünf Tagen, gelöscht werden. Weitere Informationen finden Sie unter [Angaben einer Lebenszykluskonfiguration](#).

- Aktualisierung von Log4j auf log4j 2: Wenn Sie log4j verwenden, müssen Sie aufgrund dieser Aktualisierung möglicherweise Ihre Protokollierungskonfiguration ändern. Weitere Informationen finden Sie unter [Apache log4j 2](#).

Unterschiede und Überlegungen im Bezug auf Leistung

- Leistungsunterschiede zu Tez: Mit EMR Amazon-Version 5.x ist Tez die Standard-Ausführungs-Engine für Hive anstelle von. MapReduce Tez bietet eine verbesserte Leistung für die meisten Workflows.
- Tabellen mit zahlreichen Partitionen: Abfragen, die eine große Anzahl dynamischer Partitionen generieren, schlagen möglicherweise fehl. Außerdem kann die Ausführung von Abfragen, die aus Tabellen mit vielen Partitionen auswählen, länger dauern als erwartet. Beispielsweise kann eine Auswahl aus 100.000 Partitionen 10 Minuten oder länger dauern.

Zusätzliche Funktionen von Hive bei Amazon EMR

Amazon EMR erweitert Hive um neue Funktionen, die die Hive-Integration mit anderen AWS Diensten unterstützen, z. B. die Möglichkeit, von Amazon Simple Storage Service (Amazon S3) und DynamoDB zu lesen und in diese zu schreiben.

Variablen in Hive

Sie können Variablen in Ihre Skripts aufnehmen, indem Sie das Dollarzeichen und geschweifte Klammern verwenden.

```
add jar ${LIB}/jsonserde.jar
```

Sie übergeben die Werte dieser Variablen an Hive in der Befehlszeile mit dem Parameter `-d`, wie im folgenden Beispiel dargestellt:

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

Sie können die Werte auch an Schritte übergeben, die Hive-Skripts ausführen.

So übergeben Sie Variablenwerte mithilfe der Konsole an Hive-Schritte

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Cluster erstellen.
3. Wählen Sie im Abschnitt Steps (Schritte) in Add Step (Schritt hinzufügen) die Option Hive Program (Hive-Programm) in der Liste und Configure and add (Konfigurieren und hinzufügen) aus.
4. Geben Sie im Dialogfeld Add Step (Schritt hinzufügen) die Parameter mithilfe der folgenden Tabelle als Leitfaden an. Wählen Sie anschließend Add (Hinzufügen) aus.


Feld	Aktion
Skript-S3-Speicherort*	Geben Sie an URI, wo sich Ihr Skript in Amazon S3 befindet. Der Wert muss in der folgenden Form vorliegen <i>BucketName /path/ScriptName</i> . Zum Beispiel: <code>s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q</code> .
S3-Eingabespeicherort	Geben Sie optional an, URI wo sich Ihre Eingabedateien in Amazon S3 befinden. Der Wert muss in der folgenden Form vorliegen <i>BucketName /path/</i> . Falls angegeben, wird dies als Parameter mit dem Namen <code>INPUT</code> an das Hive-Skript übergeben. Zum Beispiel: <code>s3://elasticmapreduce/samples/hive-ads/tables/</code> .
S3-Ausgabespeicherort	Geben Sie optional an, URI wo die Ausgabe in Amazon S3 gespeichert werden soll. Der Wert muss in der folgenden Form vorliegen <i>BucketName /path/</i> . Falls angegeben, wird dies als Parameter mit dem Namen <code>OUTPUT</code> an das Hive-Skript übergeben. Zum Beispiel: <code>s3://mybucket/hive-ads/output/</code> .
Argumente	Wahlweise können Sie eine Liste von Argumenten (durch Leerzeichen getrennte Zeichenfolgen) zur Übergabe an Hive eingeben. Wenn Sie in Ihrem Hive-Skript eine Pfadvariable mit dem Namen <code>\${SAMPLE}</code> definiert haben, zum Beispiel: <pre>CREATE EXTERNAL TABLE logs (requestBeginTime STRING, requestEndTime STRING, hostname STRING) PARTITIONED BY (dt STRING) \ ROW FORMAT serde 'com.amazon.elasticmapreduce.JsonSerde'</pre>

Feld	Aktion
	<pre>WITH SERDEPROPERTIES ('paths'='requestBeginTime, requestEndTime, hostname') LOCATION '\${SAMPLE}/tables/impressions';</pre> <p>Um einen Wert für die Variable zu übergeben, geben Sie im Fenster Arguments (Argumente) Folgendes ein:</p> <pre>-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/</pre>
Action on Failure	<p>Hiermit wird bestimmt, welche Schritte der Cluster als Reaktion auf Fehler ausführt. Die möglichen Werte für diese Einstellung sind folgende:</p> <ul style="list-style-type: none"> • Terminate cluster (Cluster beenden): Wenn der Schritt fehlschlägt, wird der Cluster beendet. Wenn für den Cluster der Kündigungsschutz aktiviert ist, AND Keep Alive aktiviert ist, wird er nicht beendet. • Cancel and wait (Abbrechen und warten): Wenn der Schritt fehlschlägt, werden die verbleibenden Schritte abgebrochen. Wenn für den Cluster Keepalive aktiviert ist, wird der Cluster nicht beendet. • Continue (Fortfahren): Wenn der Schritt fehlschlägt, wird mit dem nächsten Schritt fortgefahren.

5. Wählen Sie die Werte wie notwendig und anschließend Create cluster (Cluster erstellen) aus.

Um Variablenwerte an Hive zu übergeben, verwenden Sie AWS CLI

Um Variablenwerte mithilfe von an Hive-Schritte zu übergeben AWS CLI, verwenden Sie den `--steps` Parameter und fügen Sie eine Argumentliste hinzu.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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/]
```

Weitere Informationen zur Verwendung von EMR Amazon-Befehlen in der AWS CLI finden Sie unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Schritte zur Übergabe von Variablenwerten an Hive mithilfe von Java SDK

- Das folgende Beispiel zeigt, wie Variablen mithilfe von an Schritte übergeben werden. SDK Weitere Informationen finden Sie unter [Klasse StepFactory](#) in der AWS SDK for Java API Referenz.

```
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"));
```

Amazon EMR Hive-Abfragen zur Berücksichtigung partieller DynamoDB-Schemas

Amazon EMR Hive bietet maximale Flexibilität bei der Abfrage von DynamoDB-Tabellen, da Sie eine Teilmenge von Spalten angeben können, nach denen Sie Daten filtern können, anstatt zu verlangen, dass Ihre Abfrage alle Spalten umfasst. Diese Technik der Teilschemaabfrage ist effektiv, wenn Sie ein Sparse-Datenbankschema vorliegen haben und Datensätze nach wenigen Spalten filtern möchten, z. B. nach Zeitstempel.

Im folgenden Beispiel wird gezeigt, wie Sie eine Hive-Abfrage für folgende Aufgaben verwenden:

- Erstellen Sie eine DynamoDB-Tabelle.
- Wählen Sie eine Teilmenge von Elementen (Zeilen) in DynamoDB aus und grenzen Sie die Daten basierend auf bestimmten Spalten weiter ein.
- Kopieren Sie die Ergebnisdaten in 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";

```

Die folgende Tabelle zeigt die Abfragesyntax für die Auswahl einer beliebigen Kombination von Elementen aus DynamoDB.

Abfragebeispiel:	Ergebnisbeschreibung
SELECT * FROM <i>table_name</i> ;	Wählt alle Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus allen für diese Elemente verfügbaren Spalten ein.
SELECT * FROM <i>table_name</i> WHERE <i>field_name</i> = <i>value</i> ;	Wählt einige Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus allen für diese Elemente verfügbaren Spalten ein.
SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> ;	Wählt alle Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus einigen für diese Elemente verfügbaren Spalten ein.
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> ;	Wählt einige Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus einigen für diese Elemente verfügbaren Spalten ein.

Kopieren von Daten zwischen DynamoDB-Tabellen in verschiedenen AWS -Regionen

Amazon EMR Hive bietet eine `dynamodb.region` Eigenschaft, die Sie pro DynamoDB-Tabelle festlegen können. Wenn `dynamodb.region` in zwei Tabellen unterschiedlich festgelegt ist, erscheinen alle Daten, die Sie zwischen den Tabellen kopieren, automatisch in den angegebenen Regionen.

Das folgende Beispiel zeigt das Erstellen einer DynamoDB-Tabelle mit einem Hive-Skript, das die `dynamodb.region`-Eigenschaft festlegt:

Note

Die tabellenspezifischen Regionseigenschaften überschreiben die globalen Hive-Eigenschaften.

```
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");
```

Festlegen von DynamoDB-Durchsatzwerten pro Tabelle

Amazon EMR Hive ermöglicht es Ihnen, DynamoDB `readThroughputPercent` und `writeThroughputPercent` Einstellungen pro Tabelle in der Tabellendefinition festzulegen. Das folgende Amazon EMR Hive-Skript zeigt, wie die Durchsatzwerte festgelegt werden. Weitere Informationen zu DynamoDB-Durchsatzwerten finden Sie unter [Angeben von Lese- und Schreib Anforderungen für Tabellen](#).

```
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");
```

Konfigurieren eines externen Metastores für Hive

Standardmäßig zeichnet Hive Metastore-Informationen in einer Meine SQL Datenbank im Dateisystem des Primärknotens auf. Die Metastore enthält eine Beschreibung der Tabelle und der zugrunde liegenden Daten, auf denen sie beruht. Das schließt auch die Partitionsnamen, Datentypen usw. ein. Wenn ein Cluster beendet wird, werden alle Cluster-Knoten heruntergefahren – einschließlich Primärknoten. Wenn dies geschieht, gehen lokale Daten verloren, da die Dateisysteme der Knoten flüchtigen Speicher verwenden. Wenn Sie der Metastore dauerhaft sein soll, müssen Sie einen externen Metastore außerhalb des Clusters erstellen.

Ihnen stehen zwei Optionen für einen externen Metastore zur Verfügung:

- AWS Glue Data Catalog (nur EMR Amazon-Version 5.8.0 oder höher).

Weitere Informationen finden Sie unter [Den AWS Glue-Datenkatalog als Metastore für Hive verwenden](#).

- Amazon RDS oder Amazon Aurora.

Weitere Informationen finden Sie unter [Verwenden einer externen My SQL Database oder Amazon Aurora](#).

Note

Wenn Sie Hive 3 verwenden und zu viele Verbindungen zu Hive Metastore haben, konfigurieren Sie den Parameter `datanucleus.connectionPool.maxPoolSize` so, dass er einen kleineren Wert hat, oder erhöhen Sie die Anzahl der Verbindungen, die der Datenbankserver verarbeiten kann. Die erhöhte Anzahl von Verbindungen ist darauf zurückzuführen, wie Hive die maximale Anzahl von JDBC Verbindungen berechnet. [Informationen zur Berechnung des optimalen Leistungswerts finden Sie unter Hive-Konfigurationseigenschaften](#).

Den AWS Glue-Datenkatalog als Metastore für Hive verwenden

Mit EMR Amazon-Version 5.8.0 oder höher können Sie Hive so konfigurieren, dass der AWS Glue-Datenkatalog als Metastore verwendet wird. Wir empfehlen diese Konfiguration, wenn Sie einen

persistente Metastore oder einen gemeinsamen Metastore für verschiedene Cluster, Services, Anwendungen oder AWS -Konten benötigen.

AWS Glue ist ein vollständig verwalteter Service zum Extrahieren, Transformieren und Laden (ETL), mit dem Sie Ihre Daten einfach und kostengünstig kategorisieren, bereinigen, anreichern und zuverlässig zwischen verschiedenen Datenspeichern verschieben können. Der AWS Glue Data Catalog bietet ein einheitliches Metadaten-Repository für eine Vielzahl von Datenquellen und Datenformaten, das sich sowohl in Amazon EMR als auch in Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena und jede mit dem Apache Hive Metastore kompatible Anwendung integrieren lässt. AWS Glue-Crawler können automatisch Schemas aus Quelldaten in Amazon S3 ableiten und die zugehörigen Metadaten im Datenkatalog speichern. Weitere Informationen zum Datenkatalog finden Sie unter [Füllen des AWS Glue-Datenkatalogs im AWS Glue-Entwicklerhandbuch](#).

Für AWS Glue fallen separate Gebühren an. Es gibt eine monatliche Gebühr für das Speichern und Zugreifen auf die Metadaten im Datenkatalog, einen Stundensatz, der pro Minute für AWS ETL Glue-Jobs und Crawler-Laufzeit abgerechnet wird, und einen Stundensatz, der pro Minute für jeden bereitgestellten Entwicklungsendpunkt abgerechnet wird. Der Datenkatalog ermöglicht Ihnen die gebührenfreie Speicherung von bis zu einer Million Objekten. Wenn Sie mehr als eine Million Objekte speichern, wird Ihnen USD 1 USD für jeweils 100.000 Objekte über eine Million berechnet. Im Data Catalog ist ein Objekt eine Tabelle, eine Partition oder eine Datenbank. Weitere Informationen finden Sie unter [Glue-Preise](#).

Important

Wenn Sie vor dem 14. August 2017 Tabellen mit Amazon Athena oder Amazon Redshift Spectrum erstellt haben, werden Datenbanken und Tabellen in einem von Athena verwalteten Katalog gespeichert, der vom Glue-Datenkatalog getrennt ist. AWS Um Amazon EMR in diese Tabellen zu integrieren, müssen Sie ein Upgrade auf den AWS Glue Data Catalog durchführen. Weitere Informationen finden Sie unter [Upgrade auf den AWS Glue-Datenkatalog](#) im Amazon Athena Athena-Benutzerhandbuch.

AWS Glue Data Catalog als Metastore angeben

Sie können den AWS Glue-Datenkatalog als Metastore angeben, indem Sie Amazon AWS Management Console EMR API, AWS CLI, oder verwenden. Wenn Sie das CLI oder verwenden API, verwenden Sie die Konfigurationsklassifizierung für Hive, um den Datenkatalog anzugeben. Darüber

hinaus können Sie mit Amazon EMR 5.16.0 und höher die Konfigurationsklassifizierung verwenden, um einen Datenkatalog in einem anderen zu spezifizieren. AWS-Konto Wenn Sie die Konsole verwenden, können Sie den Data Catalog mit den erweiterten Optionen oder den Schnelloptionen angeben.

Console

Um AWS Glue Data Catalog mit der Konsole als Hive-Metastore anzugeben

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie EC2 im linken Navigationsbereich unter EMR on die Option Clusters und anschließend Create cluster aus.
3. Wählen Sie unter Anwendungspaket die Option Core Hadoop oder Benutzerdefiniert aus. HBase Wenn Sie Ihren Cluster anpassen, stellen Sie sicher, dass Sie Hive oder HCatalog als eine Ihrer Anwendungen auswählen.
4. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Hive-Tabellenmetadaten verwenden.
5. Wählen Sie alle anderen Optionen aus, die für Ihren Cluster gelten.
6. Um Ihren Cluster jetzt zu starten, wählen Sie Cluster erstellen aus.

CLI

Um den AWS Glue-Datenkatalog als Hive-Metastore anzugeben mit dem AWS CLI

Weitere Hinweise zur Angabe einer Konfigurationsklassifizierung mithilfe von AWS CLI und finden Sie EMR API unter [Anwendungen konfigurieren](#)

- Geben Sie den Wert für `hive.metastore.client.factory.class` mittels der Konfigurationsklassifizierung `hive-site` wie im folgenden Beispiel gezeigt an:

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

```
]
```

Wenn Sie in den EMR Release-Versionen 5.28.0, 5.28.1, 5.29.0 oder 6.x einen Cluster mit dem AWS Glue-Datenkatalog als Metastore erstellen, setzen Sie den Wert auf `hive.metastore.schema.validation false`. Dadurch wird verhindert, dass Hive und das Metastore-Schema anhand von My validieren. HCatalog SQL Ohne diese Konfiguration wird die primäre Instanzgruppe nach der Neukonfiguration auf Hive oder gesperrt. HCatalog

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.validation": "false"
    }
  }
]
```

Wenn Sie bereits über einen Cluster mit der EMR Release-Version 5.28.0, 5.28.1 oder 5.29.0 verfügen, können Sie die primäre Instanzgruppe mit den folgenden Informationen auf einrichten: `hive.metastore.schema.validation false`

```
Classification = hive-site
Property       = hive.metastore.schema.validation
Value         = false
```

Um einen Datenkatalog in einem anderen AWS Konto anzugeben, fügen Sie die `hive.metastore.glue.catalogid` Eigenschaft hinzu, wie im folgenden Beispiel gezeigt. Ersetzen Sie *acct-id* durch das AWS -Konto des Data Catalog.

```
[
  {
    "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"  
  }  
}  
]
```

IAMBerechtigungen

Das EC2 Instanzprofil für einen Cluster muss über IAM Berechtigungen für AWS Glue-Aktionen verfügen. Wenn Sie die Verschlüsselung für AWS Glue Data Catalog-Objekte aktivieren, muss die Rolle außerdem berechtigt sein, die für die Verschlüsselung AWS KMS key verwendeten Daten zu verschlüsseln, zu entschlüsseln und zu generieren.

Berechtigungen für AWS Glue-Aktionen

Wenn Sie das EC2 Standard-Instance-Profil für Amazon verwenden, ist keine Aktion erforderlich. Die `AmazonElasticMapReduceforEC2Role` verwaltete Richtlinie, die an die angehängt ist, `EMR_EC2_DefaultRole` ermöglicht alle erforderlichen AWS Glue-Aktionen. Wenn Sie jedoch ein benutzerdefiniertes EC2 Instanzprofil und Berechtigungen angeben, müssen Sie die entsprechenden AWS Glue-Aktionen konfigurieren. Verwenden Sie hierzu die verwaltete Richtlinie `AmazonElasticMapReduceforEC2Role` als Ausgangspunkt. Weitere Informationen finden Sie unter [Servicerolle für EC2 Cluster-Instances \(EC2Instance-Profil\)](#) im Amazon EMR Management Guide.

Berechtigungen zum Verschlüsseln und Entschlüsseln von AWS Glue Data Catalog

Ihr Instance-Profil benötigt die Erlaubnis, Daten mithilfe Ihres Schlüssels zu verschlüsseln und zu entschlüsseln. Sie müssen diese Berechtigungen nicht konfigurieren, wenn die beiden folgenden Aussagen zutreffen:

- Sie aktivieren die Verschlüsselung für AWS Glue-Datenkatalogobjekte mithilfe verwalteter Schlüssel für AWS Glue.
- Sie verwenden einen Cluster, der sich im selben Verzeichnis befindet AWS-Konto wie der AWS Glue-Datenkatalog.

Andernfalls müssen Sie der Berechtigungsrichtlinie, die Ihrem EC2 Instanzprofil beigefügt ist, die folgende Erklärung hinzufügen.

```
[
  {
    "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"
      }
    ]
  }
]
```

Weitere Informationen zur Verschlüsselung von AWS Glue Data Catalog finden Sie unter [Verschlüsselung Ihres Datenkatalogs](#) im AWS Glue Developer Guide.

Ressourcenbasierte Berechtigungen

Wenn Sie AWS Glue in Verbindung mit Hive, Spark oder Presto in Amazon EMR verwenden, unterstützt AWS Glue ressourcenbasierte Richtlinien zur Steuerung des Zugriffs auf Datenkatalogressourcen. Zu diesen Ressourcen gehören Datenbanken, Tabellen, Verbindungen und benutzerdefinierte Funktionen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für AWS Glue](#) im AWS -Glue-Entwicklerhandbuch.

Wenn Sie ressourcenbasierte Richtlinien verwenden, um den Zugriff auf AWS Glue von Amazon aus zu beschränken, muss der Principal, den Sie in der Berechtigungsrichtlinie angeben, die Rolle sein, die dem EC2 Instance-Profil ARN zugeordnet ist, das bei der Erstellung eines Clusters angegeben wird. Beispielsweise können Sie für eine ressourcenbasierte Richtlinie, die an einen Katalog angehängt ist, die Rolle ARN für die Standard-Service-Rolle für Cluster-Instances angeben. EC2 *EMR_EC2_DefaultRole* als der Principal, wobei das im folgenden Beispiel gezeigte Format verwendet wird:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Das Tool *acct-id* kann sich von der AWS Glue-Konto-ID unterscheiden. Dies ermöglicht den Zugriff von EMR Clustern in verschiedenen Konten aus. Sie können mehrere Principals angeben, von denen jeder aus einem anderen Konto stammt.

Überlegungen zur Verwendung von AWS Glue Data Catalog

Beachten Sie die folgenden Punkte, wenn Sie den AWS Glue-Datenkatalog als Metastore mit Hive verwenden:

- Das Hinzufügen von JARs Hilfsmitteln mithilfe der Hive-Shell wird nicht unterstützt. Um das Problem zu umgehen, verwenden Sie die `hive-site` Konfigurationsklassifizierung, um die `hive.aux.jars.path` Eigenschaft festzulegen, wodurch dem Hive-Klassenpfad JARs ein Hilfsprogramm hinzugefügt wird.
- [Hive-Transaktionen](#) werden nicht unterstützt.
- Das Umbenennen von Tabellen in AWS Glue wird nicht unterstützt.
- Wenn Sie eine Hive-Tabelle erstellen, ohne eine `LOCATION` anzugeben, werden die Tabellendaten an dem in der Eigenschaft `hive.metastore.warehouse.dir` angegebenen Ort gespeichert. Standardmäßig ist dies ein Speicherort in HDFS. Wenn ein anderer Cluster auf die Tabelle zugreifen muss, schlägt er fehl, sofern er nicht über ausreichende Berechtigungen für den Cluster verfügt, der die Tabelle erstellt hat. Da der HDFS Speicher vorübergehend ist, gehen außerdem die Tabellendaten verloren, wenn der Cluster beendet wird, und die Tabelle muss neu erstellt werden. Wir empfehlen, dass Sie `LOCATION` in Amazon S3 angeben, wenn Sie eine Hive-Tabelle mit AWS Glue erstellen. Alternativ können Sie die `hive-site`-Konfigurationsklassifizierung verwenden, um einen Speicherort in Amazon S3 für `hive.metastore.warehouse.dir` anzugeben, der für alle Hive-Tabellen gilt. Wenn eine Tabelle an einem HDFS Standort erstellt wurde und der Cluster, der sie erstellt hat, noch läuft, können Sie den Tabellenstandort von AWS Glue aus auf Amazon S3 aktualisieren. Weitere Informationen finden Sie unter [Arbeiten mit Tabellen auf der AWS Glue-Konsole](#) im AWS Glue-Entwicklerhandbuch.
- Partitionswerte, die Anführungszeichen und Apostrophe enthalten, werden nicht unterstützt, zum Beispiel `PARTITION (owner="Doe 's")`.
- [Spaltenstatistiken](#) werden für emr-5.31.0 und höher unterstützt.
- Die Verwendung der [Hive-Autorisierung](#) wird nicht unterstützt. Als Alternative sollten Sie die Verwendung [ressourcenbasierter Richtlinien für AWS Glue](#) in Betracht ziehen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für den Amazon EMR Access to AWS Glue-Datenkatalog](#).
- [Hive-Einschränkungen](#) werden nicht unterstützt.

- Die [kostenbasierte Optimierung in Hive](#) wird nicht unterstützt.
- Die Einstellung `hive.metastore.partition.inherit.table.properties` wird nicht unterstützt.
- Die Verwendung der folgenden Metastore-Konstanten wird nicht unterstützt:
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.
- Wenn Sie einen Prädikatausdruck verwenden, müssen sich explizite Werte auf der rechten Seite des Vergleichsoperators befinden. Andernfalls schlagen Abfragen möglicherweise fehl.
 - Richtig: `SELECT * FROM mytable WHERE time > 11`
 - Falsch: `SELECT * FROM mytable WHERE 11 > time`
- Die EMR Amazon-Versionen 5.32.0 und 6.3.0 und höher unterstützen die Verwendung benutzerdefinierter Funktionen (UDFs) in Prädikatausdrücken. Wenn Sie frühere Versionen verwenden, schlagen Ihre Abfragen möglicherweise fehl, weil Hive versucht, die Abfrageausführung zu optimieren.
- [Temporäre Tabellen](#) werden nicht unterstützt.
- Wir empfehlen, Tabellen mit Anwendungen über Amazon zu erstellen, EMR anstatt sie direkt mit AWS Glue zu erstellen. Das Erstellen einer Tabelle mit AWS Glue kann dazu führen, dass erforderliche Felder fehlen und Abfrageausnahmen auftreten.
- In Version EMR 5.20.0 oder höher wird die parallel Partitionsbereinigung automatisch für Spark und Hive aktiviert, wenn AWS Glue Data Catalog als Metastore verwendet wird. Diese Änderung reduziert die Zeit für die Abfrageplanung erheblich, da mehrere Anfragen parallel ausgeführt werden, um Partitionen abzurufen. Die Gesamtzahl der Segmente, die gleichzeitig ausgeführt werden können, liegt zwischen 1 und 10. Der Standardwert ist 5, was eine empfohlene Einstellung ist. Sie können dies ändern, indem Sie die Eigenschaft `aws.glue.partition.num.segments` in der `hive-site`-Konfigurationsklassifizierung angeben. Wenn eine Drosselung auftritt, können Sie das Feature ausschalten, indem Sie den Wert auf 1 ändern. Weitere Informationen finden Sie unter [AWS -Glue-Segmentstruktur](#).

Verwenden einer externen My SQL Database oder Amazon Aurora

Um eine externe SQL My-Datenbank oder Amazon Aurora als Ihren Hive-Metastore zu verwenden, überschreiben Sie die Standardkonfigurationswerte für den Metastore in Hive, um den externen Datenbankspeicherort anzugeben, entweder auf einer Amazon RDS SQL My-Instance oder einem Amazon Aurora PostgreSQL Instance

Note

Von Hive werden gleichzeitige Schreibzugriffe auf Metastore-Tabellen weder unterstützt noch verhindert. Wenn zwei Cluster gemeinsam Metastore-Daten nutzen, müssen Sie sicherstellen, dass Sie nicht gleichzeitig in dieselbe Metastore-Tabelle schreiben, es sei denn, sie schreiben auf verschiedene Partitionen derselben Metastore-Tabelle.

Das folgende Verfahren zeigt, wie Sie die Standardkonfigurationswerte für den Speicherort des Hive-Metastores überschreiben und einen Cluster mit dem neu konfigurierten Metastore-Speicherort starten.

Um einen Metastore zu erstellen, der sich außerhalb des Clusters befindet EMR

1. Erstellen Sie eine My SQL - oder Aurora SQL PostgreSQL-Datenbank. Wenn Sie PostgreSQL verwenden, müssen Sie es konfigurieren, nachdem Sie Ihren Cluster bereitgestellt haben. Nur My SQL wird bei der Clustererstellung unterstützt. Informationen zu den Unterschieden zwischen Aurora My SQL und Aurora PostgreSQL finden Sie unter [Überblick über Amazon Aurora My SQL](#) und [Arbeiten mit Amazon Aurora SQL PostgreSQL](#). Allgemeine Informationen zum Erstellen einer RDS Amazon-Datenbank finden Sie unter <https://aws.amazon.com/rds/>.
2. Ändern Sie Ihre Sicherheitsgruppen, um JDBC Verbindungen zwischen Ihrer Datenbank und der ElasticMapReduce-Master-Sicherheitsgruppe zuzulassen. Informationen dazu, wie Sie Ihre Sicherheitsgruppen für den Zugriff ändern können, finden Sie unter [Arbeiten mit von Amazon EMR verwalteten Sicherheitsgruppen](#).
3. Legen Sie die JDBC Konfigurationswerte fest in `hive-site.xml`:

Important

Wenn Sie vertrauliche Informationen wie Passwörter für die EMR Amazon-Konfiguration angeben, werden diese Informationen für die Konten angezeigt, die über ausreichende Berechtigungen verfügen. Wenn Sie befürchten, dass diese

Informationen anderen Benutzern angezeigt werden könnten, erstellen Sie den Cluster mit einem Administratorkonto und beschränken Sie andere Benutzer (Benutzer oder IAM Benutzer mit delegierten Anmeldeinformationen) auf den Zugriff auf Dienste im Cluster, indem Sie eine Rolle erstellen, die explizit Berechtigungen für den `elasticmapreduce:DescribeCluster` API Schlüssel verweigert.

- a. Erstellen Sie eine Konfigurationsdatei namens `hiveConfiguration.json`, die Änderungen an `hive-site.xml` enthält, wie im folgenden Beispiel gezeigt.

Ersetzen *hostname* mit der DNS Adresse Ihrer RDS Amazon-Instance, auf der die Datenbank läuft, und *username* and *password* mit den Anmeldeinformationen für Ihre Datenbank. Weitere Informationen zum Herstellen einer Verbindung zu My SQL - und Aurora-Datenbank-Instances finden Sie unter [Verbindung zu einer DB-Instance herstellen, auf der die My SQL Database-Engine ausgeführt wird](#), und [Herstellen einer Verbindung zu einem Athena-DB-Cluster](#) im RDS Amazon-Benutzerhandbuch. `javax.jdo.option.ConnectionURL` ist die JDBC Verbindungszeichenfolge für einen JDBC Metastore. `javax.jdo.option.ConnectionDriverName` ist der Treiberklassenname für einen JDBC Metastore.


Die Meine SQL JDBC Treiber werden von Amazon installiert EMR.

Die Werteigenschaft darf keine Leerzeichen oder Wagenrückläufe enthalten. Sie sollte in einer Zeile erscheinen.

```
[
  {
    "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. Verweisen Sie bei der Erstellung des Clusters auf die `hiveConfiguration.json` Datei, wie im folgenden AWS CLI Befehl gezeigt. In diesem Befehl ist die Datei lokal gespeichert.

Sie können die Datei auch zu Amazon S3 hochladen und dort referenzieren, beispielsweise `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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. Stellen Sie eine Verbindung mit dem Primärknoten Ihres Clusters her.

Informationen zum Herstellen einer Connect dem primären Knoten finden Sie unter [Verbindung zum primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.

5. Erstellen Sie Ihre Hive-Tabellen, indem Sie den Speicherort in Amazon S3 durch Eingabe eines Befehls ähnlich dem folgenden angeben:

```
CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
key int,
value int
)
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. Fügen Sie dem aktiven Cluster Ihr Hive-Skript hinzu.

Ihr Hive-Cluster wird mit dem Metastore in Amazon ausgeführt. RDS Starten Sie alle zusätzlichen Hive-Cluster, die diesen Metastore nutzen, indem Sie den Metastore-Speicherort angeben.

Den Hive-JDBC-Treiber verwenden

Sie können beliebige Business-Intelligence-Tools wie Microsoft Excel, MicroStrategy, QlikView und Tableau mit Amazon EMR verwenden, um Daten zu erkunden und zu visualisieren. Viele dieser

Tools erfordern einen Java Database Connectivity (JDBC)- oder einen Open Database Connectivity (ODBC)-Treiber. Amazon EMR unterstützt JDBC- und ODBC-Verbindungen.

Das folgende Beispiel zeigt die Verwendung von SQL Workbench/J als SQL-Client zum Verbinden eines Hive-Clusters in Amazon EMR. Weitere Treiber finden Sie unter [Verwenden von Business-Intelligence-Tools mit Amazon EMR](#).

Bevor Sie SQL Workbench/J installieren und damit arbeiten, laden Sie das Treiber-Paket herunter, und installieren Sie den Treiber. Die in dem Paket enthaltenen Treiber unterstützen die Hive-Versionen, die in den Amazon-EMR-Versionen 4.0 und höher verfügbar sind. Detaillierte Versionshinweise und Dokumentationen finden Sie in der PDF-Dokumentation in dem Paket.

- Das neueste Hive-JDBC-Treiberpaket herunterladen

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/>

- Ältere Versionen des Hive-JDBC-Treibers

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/>

So installieren und konfigurieren Sie die SQL Workbench

1. Laden Sie über <http://www.sql-workbench.net/downloads.html> den SQL Workbench/J-Client für Ihr Betriebssystem herunter.
2. Installieren Sie SQL Workbench/J. Weitere Informationen finden Sie unter [Installieren und Starten von SQL Workbench/J](#) im Benutzerhandbuch für SQL Workbench/J.
3. Linux-, Unix-, Mac OS X-Benutzer: Erstellen Sie mithilfe des folgenden Befehls in einer Terminalsitzung einen SSH-Tunnel zum Master-Knoten Ihres Clusters. Ersetzen Sie *master-public-dns-name* durch den öffentlichen DNS-Namen des Hauptknotens und *path-to-key-file* mit dem Speicherort und dem Dateinamen Ihrer Datei für den privaten Amazon-EC2-Schlüssel (.pem).

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000  
hadoop@master-public-dns-name
```

Windows-Benutzer: Erstellen Sie in einer PuTTY-Sitzung einen SSH-Tunnel zum Hauptknoten Ihres Clusters (mit lokaler Port-Weiterleitung) und legen Sie 10000 für Quellport und *master-public-dns-name*:10000 für Ziel fest. Ersetzen Sie *master-public-dns-name* durch den öffentlichen DNS-Namen des Master-Knotens.

4. Fügen Sie SQL Workbench den JDBC-Treiber hinzu.
 - a. Klicken Sie im Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) auf Manage Drivers (Treiber verwalten).
 - b. Klicken Sie auf das Symbol (leere Seite) Create a new entry (Neuen Eintrag erstellen).
 - c. Geben Sie in das Feld Name **Hive JDBC** ein.
 - d. Klicken Sie bei Library (Bibliothek) auf das Symbol Select the JAR file(s) (JAR-Datei(en) auswählen).
 - e. Navigieren Sie zu dem Speicherort, der die extrahierten Treiber enthält. Wählen Sie die Treiber aus, die in der heruntergeladenen Version des JDBC-Treiberpakets enthalten sind, und klicken Sie auf Öffnen.

Ihr JDBC-Treiberpaket kann beispielsweise die folgenden JAR-Dateien enthalten.

```
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. Wählen Sie im Dialogfeld Please select one driver (Einen Treiber auswählen) die Optionen `com.amazon.hive.jdbc41.HS2Driver` und OK (OK) aus.
5. Wenn Sie zum Dialogfeld Manage Drivers (Treiber verwalten) zurückkehren, überprüfen Sie, ob das Feld Classname (Klassenname) ausgefüllt ist. Klicken Sie dann auf OK (OK).
6. Wenn Sie zum Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) zurückkehren, überprüfen Sie, ob das Dialogfeld Driver (Treiber) auf Hive JDBC (Hive JDBC) festgelegt ist. Geben Sie die folgende JDBC-Verbindungszeichenfolge im Feld URL (URL) ein:
`jdbc:hive2://localhost:10000/default`.
7. Wählen Sie OK (OK) aus, um die Verbindung herzustellen. Nachdem die Verbindung aufgebaut ist, werden oben im SQL Workbench/J-Fenster Verbindungsdetails angezeigt.

Weitere Informationen zum Verwenden von Hive und der JDBC-Schnittstelle finden Sie unter [HiveClient](#) und [HiveJDBCInterface](#) in der Apache Hive-Dokumentation.

Verbessern der Hive-Leistung

Amazon EMR bietet Funktionen zur Leistungsoptimierung bei der Verwendung von Hive zum Abfragen, Lesen und Schreiben von in Amazon S3 gespeicherten Daten.

S3 Select kann die Abfrageleistung für CSV und JSON Dateien in einigen Anwendungen verbessern, indem die Verarbeitung auf Amazon S3 „heruntergeschoben“ wird.

Der EMRFS S3-optimierte Committer ist eine Alternative zu dieser [OutputCommitter](#)-Klasse, die Auflisten- und Umbenennungsvorgänge überflüssig macht, um die Leistung beim Schreiben von Dateien zu verbessern, die Amazon S3 verwendet.

Themen

- [Aktivierung des optimierten Hive EMRFS S3-Committers](#)
- [Verwendung von S3 Select mit Hive zur Verbesserung der Leistung](#)
- [MSCKOptimierung](#)

Aktivierung des optimierten Hive EMRFS S3-Committers

Der Hive EMRFS S3 Optimized Committer ist eine alternative Methode, mit der EMR Hive bei der Verwendung von Dateien für Insert-Abfragen schreibt. EMRFS Der Committer macht Listen- und Umbenennungsvorgänge überflüssig, die in Amazon S3 ausgeführt wurden, und verbessert die Leistung der Anwendung. Die Funktion ist ab Version EMR 5.34 und 6.5 verfügbar. EMR

Aktivieren des Committers

Wenn Sie EMR Hive so einrichten möchten, dass Daten als Standard für alle von Hive verwalteten und externen Tabellen übertragen werden, verwenden Sie die folgende `hive-site` Konfiguration in Clustern der Versionen EMR 6.5.0 oder 5.34.0. `HiveEMRFSOptimizedCommitter` EMR

```
[
  {
    "classification": "hive-site",
    "properties": {
```

```
        "hive.blobstore.use.output-committer": "true"
    }
}
]
```

Note

Schalten Sie dieses Feature nicht ein, wenn `hive.exec.parallel` auf `true` eingestellt ist.

Einschränkungen

Die folgenden grundlegenden Einschränkungen gelten für Tags (Markierungen):

- Die Aktivierung von Hive für das automatische Zusammenführen kleiner Dateien wird nicht unterstützt. Die standardmäßige Hive-Commit-Logik wird auch dann verwendet, wenn der optimierte Committer aktiviert ist.
- Hive-Tabellen ACID werden nicht unterstützt. Die standardmäßige Hive-Commit-Logik wird auch dann verwendet, wenn der optimierte Committer aktiviert ist.
- Die Benennungsnomenklatur für geschriebene Dateien wurde von Hives von `<task_id>_<attempt_id>_<copy_n>` auf `<task_id>_<attempt_id>_<copy_n>_<query_id>` geändert. Zum Beispiel eine Datei mit dem Namen `s3://warehouse/table/partition=1/000000_0` wird geändert zu `s3://warehouse/table/partition=1/000000_0-hadoop_20210714130459_ba7c23ec-5695-4947-9d98-8a40ef759222-1`. Das `query_id` hier ist eine Kombination aus Benutzername, Zeitstempel und UUID.
- Wenn sich benutzerdefinierte Partitionen auf verschiedenen Dateisystemen (HDFS, S3) befinden, wird diese Funktion automatisch deaktiviert. Wenn sie aktiviert ist, wird die standardmäßige Hive-Commit-Logik verwendet.

Verwendung von S3 Select mit Hive zur Verbesserung der Leistung

Mit der EMR Amazon-Release-Version 5.18.0 und höher können Sie [S3 Select](#) mit Hive auf Amazon verwenden. EMR S3 Select ermöglicht es Anwendungen, nur eine Teilmenge von Daten aus einem Objekt abzurufen. Für Amazon EMR wird die Rechenarbeit beim Filtern großer Datensätze für

die Verarbeitung vom Cluster auf Amazon S3 „heruntergeschoben“, wodurch die Leistung einiger Anwendungen verbessert und die zwischen Amazon EMR und Amazon S3 übertragene Datenmenge reduziert werden kann.

S3 Select wird mit Hive-Tabellen unterstützt, die auf CSV und JSON -Dateien basieren, und indem Sie die `s3select.filter` Konfigurationsvariable `true` während Ihrer Hive-Sitzung auf setzen. Weitere Informationen und Beispiele finden Sie unter [Angeben von S3 Select in Ihrem Code](#).

Ist S3 Select das Richtige für meine Anwendung?

Wir empfehlen, dass Sie Benchmark-Tests für Ihre Anwendungen im Vergleich mit und ohne S3 Select durchführen, um zu sehen, ob es für Ihre Anwendung geeignet sein könnte.

Verwenden Sie die folgenden Richtlinien, um zu bestimmen, ob Ihre Anwendung ein Kandidat für die Verwendung von S3 ist:

- Ihre Abfrage filtert mehr als die Hälfte des ursprünglichen Datensatzes.
- Ihre Abfrage-Filterprädikate verwenden Spalten mit einem Datentyp, der von Amazon S3 Select unterstützt wird. Weitere Informationen finden Sie unter [Datentypen](#) im Benutzerhandbuch für Amazon Simple Storage Service.
- Ihre Netzwerkverbindung zwischen Amazon S3 und dem EMR Amazon-Cluster hat eine gute Übertragungsgeschwindigkeit und verfügbare Bandbreite. Amazon S3 komprimiert HTTP Antworten nicht, daher wird die Antwortgröße bei komprimierten Eingabedateien wahrscheinlich zunehmen.

Überlegungen und Einschränkungen

- Serverseitige Amazon S3 S3-Verschlüsselung mit vom Kunden bereitgestellten Verschlüsselungsschlüsseln (SSE-C) und clientseitige Verschlüsselung werden nicht unterstützt.
- Die Eigenschaft `AllowQuotedRecordDelimiters` wird nicht unterstützt. Wenn diese Eigenschaft angegeben ist, schlägt die Abfrage fehl.
- Es werden nur JSON Dateien im CSV -8-Format unterstützt UTF. Mehrzeilig CSVs und JSON werden nicht unterstützt.
- Nur unkomprimierte oder gzip- oder bzip2-Dateien werden unterstützt.
- Kommentarzeichen auf der letzten Zeile werden nicht unterstützt.
- Leere Zeilen am Ende einer Datei werden nicht verarbeitet.

- Hive on Amazon EMR unterstützt die primitiven Datentypen, die S3 Select unterstützt. Weitere Informationen finden Sie unter [Datentypen](#) im Benutzerhandbuch für Amazon Simple Storage Service.

Angaben von S3 Select in Ihrem Code

Um S3 Select in Ihrer Hive-Tabelle zu verwenden, erstellen Sie die Tabelle durch Angabe von `com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` als `INPUTFORMAT`-Klassenname, und geben Sie einen Wert für die Eigenschaft `s3select.format` mit der `TBLPROPERTIES`-Klausel an.

S3 Select ist standardmäßig deaktiviert, wenn Sie Abfragen ausführen. Aktivieren Sie S3 Select, indem Sie `s3select.filter` auf `true` in Ihrer Hive-Sitzung setzen, wie unten gezeigt. Die folgenden Beispiele zeigen, wie Sie S3 Select angeben, wenn Sie eine Tabelle aus JSON Basisdateien erstellen CSV und die Tabelle anschließend mit einer einfachen Select-Anweisung abfragen.

Example CREATETABLEAnweisung für eine CSV basierte Tabelle

```
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.q1.io.HiveIgnoreKeyTextOutputFormat'  
LOCATION 's3://path/to/mycsvfile/'  
TBLPROPERTIES (  
  "s3select.format" = "csv",  
  "s3select.headerInfo" = "ignore"  
);
```

Example CREATETABLEAussage für eine JSON basierte Tabelle

```
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.q1.io.HiveIgnoreKeyTextOutputFormat'  
LOCATION 's3://path/to/json/'  
TBLPROPERTIES (  
  "s3select.format" = "json"  
);
```

Example SELECTTABLEAussage

```
SET s3select.filter=true;  
SELECT * FROM mys3selecttable WHERE col2 > 10;
```

MSCKOptimierung

Hive speichert eine Liste von Partitionen für jede Tabelle in seinem Metastore. Wenn Partitionen jedoch direkt zum Dateisystem hinzugefügt oder daraus entfernt werden, bemerkt der Hive-Metastore diese Änderungen nicht. Der [MSCKBefehl](#) aktualisiert die Partitionsmetadaten im Hive-Metastore für Partitionen, die direkt zum Dateisystem hinzugefügt oder daraus entfernt wurden. Die Syntax für den Befehl lautet:

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```

Hive implementiert diesen Befehl wie folgt:

1. Hive ruft alle Partitionen für die Tabelle aus dem Metastore ab. Aus der Liste der Partitionspfade, die im Dateisystem nicht existieren, wird dann eine Liste von Partitionen erstellt, die aus dem Metastore gelöscht werden sollen.
2. Hive sammelt die im Dateisystem vorhandenen Partitionspfade, vergleicht sie mit der Liste der Partitionen aus dem Metastore und generiert eine Liste von Partitionen, die dem Metastore hinzugefügt werden müssen.
3. Hive aktualisiert den Metastore im Modus ADD, DROP oder SYNC.

Note

Wenn der Metastore viele Partitionen enthält, dauert die Ausführung des Schritts zur Überprüfung, ob eine Partition nicht im Dateisystem existiert, sehr lange, da der exists API Aufruf des Dateisystems für jede Partition erfolgen muss.

In Amazon EMR 6.5.0 führte Hive eine Flagge namens ein. `hive.emr.optimize.msck.fs.check`. Wenn dieses Flag aktiviert ist, veranlasst es Hive, anhand der Liste der Partitionspfade aus dem Dateisystem, die in Schritt 2 oben generiert wurde, nach dem Vorhandensein einer Partition zu suchen, anstatt Dateisystemaufrufe durchzuführen. API In Amazon EMR 6.8.0 hat Hive diese Optimierung standardmäßig aktiviert, sodass die Markierung nicht mehr gesetzt werden muss. `hive.emr.optimize.msck.fs.check`

Verwendung von Hive Live Long und Process () LLAP

Amazon EMR 6.0.0 unterstützt die Funktionen Live Long und Process (LLAP) für Hive. LLAP verwendet persistente Daemons mit intelligentem In-Memory-Caching, um die Abfrageleistung im Vergleich zum vorherigen Standard-Tez-Container-Ausführungsmodus zu verbessern.

Die LLAP Hive-Daemons werden als Dienst verwaltet und ausgeführt. YARN Da ein YARN Dienst als lang laufende YARN Anwendung betrachtet werden kann, sind einige Ihrer Clusterressourcen Hive vorbehalten LLAP und können nicht für andere Workloads verwendet werden. [Weitere Informationen finden Sie unter LLAPService. YARN API](#)

Hive LLAP auf Amazon aktivieren EMR

Um Hive LLAP auf Amazon zu aktivieren EMR, geben Sie beim Starten eines Clusters die folgende Konfiguration an.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

Weitere Informationen finden Sie unter [Konfigurieren von Anwendungen](#).

Standardmäßig weist Amazon etwa 60 Prozent der YARN Cluster-Ressourcen LLAP Hive-Daemons zu. Sie können den Prozentsatz der YARN Cluster-Ressourcen, die Hive zugewiesen sind, LLAP und die Anzahl der Task- und Core-Knoten, die bei der Hive-Zuweisung berücksichtigt werden sollen, konfigurieren. LLAP

Bei der folgenden Konfiguration wird Hive beispielsweise LLAP mit drei Daemons auf drei Aufgaben- oder Kernknoten gestartet und 40 Prozent der Ressourcen der drei Kern- oder Aufgabenknoten werden den Hive-Daemons zugewiesen. YARN LLAP

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]
```

Sie können die folgenden `hive-site` Konfigurationen in der Klassifizierung verwenden, um die Standardressourceneinstellungen zu überschreiben. API LLAP

Property (Eigenschaft)	Description (Beschreibung)
<code>hive.llap.daemon.yarn.container.mb</code>	Gesamtgröße des LLAP Daemon-Containers (in MB)
<code>hive.llap.daemon.memory.per.instance.mb</code>	Der Gesamtspeicher, der von den Executoren im LLAP Daemon-Container verwendet wird (in MB)
<code>hive.llap.io.memory.size</code>	Cache-Größe für Eingabe/Ausgabe LLAP
<code>hive.llap.daemon.num.executors</code>	Anzahl der Executoren pro Daemon LLAP

Starten Sie Hive manuell LLAP auf Ihrem Cluster

Alle Abhängigkeiten und Konfigurationen, die von verwendet werden, LLAP werden im Rahmen des Cluster-Starts in das LLAP Tar-Archiv gepackt. Wenn aktiviert LLAP ist, empfehlen wir Ihnen "hive.llap.enabled": "true", die EMR Amazon-Neukonfiguration zu verwenden, um Konfigurationsänderungen an LLAP vorzunehmen.

Andernfalls müssen Sie bei manuellen Änderungen hive-site.xml das LLAP Tar-Archiv mithilfe des `hive --service llap` Befehls neu erstellen, wie das folgende Beispiel zeigt.

```
# 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
```

Überprüfen Sie den Hive-Status LLAP

Verwenden Sie den folgenden Befehl, um den Status von Hive LLAP über Hive zu überprüfen.

```
hive --service llapstatus
```

Verwenden Sie den folgenden Befehl, um den Status von Hive LLAP using zu überprüfen. YARN

```
yarn app -status (name-of-llap-service)

# example:
```

```
yarn app -status llap0 | jq
```

Starten oder beenden Sie Hive LLAP

Da Hive als persistenter YARN Dienst LLAP ausgeführt wird, beenden oder starten Sie den YARN Dienst neu, um Hive LLAP zu beenden oder neu zu starten. Die folgenden Befehle veranschaulichen dies.

```
yarn app -stop llap0  
yarn app -start llap0
```

Ändern Sie die Größe der Anzahl der Hive-Daemons LLAP

Verwenden Sie den folgenden Befehl, um die Anzahl der LLAP Instanzen zu reduzieren.

```
yarn app -flex llap0 -component llap -1
```

Weitere Informationen finden Sie unter [Flexible auf eine Komponente eines Services reagieren](#).

Verschlüsselung im Hive

In diesem Abschnitt werden die Verschlüsselungstypen beschrieben, EMR die Amazon unterstützt.

Modulare Parquet-Verschlüsselung in Hive

Die modulare Verschlüsselung von Parquet bietet Zugriffskontrolle und Verschlüsselung auf Spaltenebene, um den Datenschutz und die Datenintegrität für Daten zu verbessern, die im Parquet-Dateiformat gespeichert sind. Diese Funktion ist in Amazon EMR Hive ab Version 6.6.0 verfügbar.

Bisher unterstützte Lösungen für Sicherheit und Integrität, die das Verschlüsseln von Dateien oder das Verschlüsseln der Speicherebene beinhalten, werden im Amazon EMR Management Guide unter [Verschlüsselungsoptionen](#) beschrieben. Diese Lösungen können für Parquet-Dateien verwendet werden, aber die Nutzung der neuen Features des integrierten Parquet-Verschlüsselungsmechanismus ermöglicht einen detaillierten Zugriff auf die Spaltenebene sowie Verbesserungen in Bezug auf Leistung und Sicherheit. Weitere Informationen zu diesem Feature finden Sie auf der Apache-Github-Seite [Parquet Modular Encryption](#).

Benutzer geben Konfigurationen mithilfe von Hadoop-Konfigurationen an Parquet-Leser und -Writer weiter. Die detaillierten Konfigurationen für Benutzer zur Konfiguration von Lesegeräten und

Schreibern zur Aktivierung der Verschlüsselung sowie zum Umschalten erweiterter Funktionen sind unter [PARQUET-1854: Eigenschaftsgesteuerte](#) Schnittstelle zu Parquet Encryption Management dokumentiert

Verwendungsbeispiele

Das folgende Beispiel behandelt das Erstellen und Schreiben in eine Hive-Tabelle mit AWS KMS , die zur Verwaltung von Verschlüsselungsschlüsseln verwendet wird.

1. Implementieren Sie eine KmsClient für den AWS KMS Dienst, wie im Dokument [PARQUET-1373: Encryption Key Management Tools](#) beschrieben. Das folgende Beispiel zeigt einen Implementierungsausschnitt.

```
package org.apache.parquet.crypto.keytools;

import com.amazonaws.AmazonClientException;
import com.amazonaws.AmazonServiceException;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.kms.AWSKMS;
import com.amazonaws.services.kms.AWSKMSClientBuilder;
import com.amazonaws.services.kms.model.DecryptRequest;
import com.amazonaws.services.kms.model.EncryptRequest;
import com.amazonaws.util.Base64;
import org.apache.hadoop.conf.Configuration;
import org.apache.parquet.crypto.KeyAccessDeniedException;
import org.apache.parquet.crypto.ParquetCryptoRuntimeException;
import org.apache.parquet.crypto.keytools.KmsClient;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

import java.nio.ByteBuffer;
import java.nio.charset.Charset;
import java.nio.charset.StandardCharsets;

public class AwsKmsClient implements KmsClient {

    private static final AWSKMS AWSKMS_CLIENT = AWSKMSClientBuilder
        .standard()
        .withRegion(Regions.US_WEST_2)
        .build();
    public static final Logger LOG = LoggerFactory.getLogger(AwsKmsClient.class);

    private String kmsToken;
```

```
private Configuration hadoopConfiguration;

@Override
public void initialize(Configuration configuration, String kmsInstanceID, String
kmsInstanceURL, String accessToken) throws KeyAccessDeniedException {
    hadoopConfiguration = configuration;
    kmsToken = accessToken;
}

@Override
public String wrapKey(byte[] keyBytes, String masterKeyIdentifier) throws
KeyAccessDeniedException {
    String value = null;
    try {
        ByteBuffer plaintext = ByteBuffer.wrap(keyBytes);

        EncryptRequest req = new
EncryptRequest().withKeyId(masterKeyIdentifier).withPlaintext(plaintext);
        ByteBuffer ciphertext = AWSKMS_CLIENT.encrypt(req).getCiphertextBlob();

        byte[] base64EncodedValue = Base64.encode(ciphertext.array());
        value = new String(base64EncodedValue, Charset.forName("UTF-8"));
    } catch (AmazonClientException ae) {
        throw new KeyAccessDeniedException(ae.getMessage());
    }
    return value;
}

@Override
public byte[] unwrapKey(String wrappedKey, String masterKeyIdentifier) throws
KeyAccessDeniedException {
    byte[] arr = null;
    try {
        ByteBuffer ciphertext =
ByteBuffer.wrap(Base64.decode(wrappedKey.getBytes(StandardCharsets.UTF_8)));
        DecryptRequest request = new
DecryptRequest().withKeyId(masterKeyIdentifier).withCiphertextBlob(ciphertext);
        ByteBuffer decipheredtext =
AWSKMS_CLIENT.decrypt(request).getPlaintext();
        arr = new byte[decipheredtext.remaining()];
        decipheredtext.get(arr);
    } catch (AmazonClientException ae) {
        throw new KeyAccessDeniedException(ae.getMessage());
    }
}
```

```

    }
    return arr;
  }
}

```

- Erstellen Sie Ihre AWS KMS Verschlüsselungsschlüssel für die Fußzeile sowie für die Spalten, auf die Ihre IAM Rollen Zugriff haben, wie unter [Schlüssel erstellen im Entwicklerhandbuch](#) beschrieben. AWS Key Management Service Die IAM Standardrolle ist EMR _ ECS _ default.
- Fügen Sie in der Hive-Anwendung auf einem EMR Amazon-Cluster den obigen Client mithilfe der ADD JAR Anweisung hinzu, wie in der [Apache Hive Resources-Dokumentation](#) beschrieben. Nachfolgend finden Sie eine Beispielanweisung:

```
ADD JAR 's3://location-to-custom-jar';
```

Eine alternative Methode besteht darin, den `auxlib` von Hive mithilfe einer Bootstrap-Aktion JAR zu erweitern. Im Folgenden finden Sie eine Beispielzeile, die der Bootstrap-Aktion hinzugefügt werden soll:

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

- Legen Sie die folgenden Konfigurationen fest:

```

set
  parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFactory;
set
  parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;

```

- Erstellen Sie eine Hive-Tabelle im Parquet-Format, geben Sie die AWS KMS Schlüssel an `SERDEPROPERTIES` und fügen Sie einige Daten ein:

```

CREATE TABLE my_table(name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
WITH SERDEPROPERTIES (
  'parquet.encryption.column.key'=<aws-kms-key-id-for-column-1>: credit_card',
  'parquet.encryption.footer.key'='<aws-kms-key-id-for-footer>')
STORED AS parquet
LOCATION "s3://<bucket/<warehouse-location>/my_table";

INSERT INTO my_table SELECT

```

```
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',5) as
  name,
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',10) as
  credit_card
from (select 1) x lateral view posexplode(split(space(100),' ')) pe as i,x;

select * from my_table;
```

6. Stellen Sie sicher, dass Sie die Daten nicht lesen können, wenn Sie am selben Speicherort eine externe Tabelle ohne Zugriff auf AWS KMS Schlüssel erstellen (z. B. IAM Rollenzugriff verweigert).

```
CREATE EXTERNAL TABLE ext_table (name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive.q1.io.parquet.serde.ParquetHiveSerDe'
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

SELECT * FROM ext_table;
```

7. Die letzte Anweisung sollte die folgende Ausnahme auslösen:

```
Failed with exception
java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
access denied
```

Verschlüsselung während der Übertragung in 2 HiveServer

Ab EMR Amazon-Version 6.9.0 ist HiveServer 2 (HS2) als Teil der [Verschlüsselung während der Übertragung in 2 HiveServer](#) Sicherheitskonfiguration TLS/SSL-aktiviert. Dies wirkt sich darauf aus, wie Sie eine Verbindung zur HS2 Ausführung auf einem EMR Amazon-Cluster herstellen, bei dem die Verschlüsselung während der Übertragung aktiviert ist. Um eine Verbindung herzustellen HS2, müssen Sie die TRUSTSTORE_PASSWORD Parameterwerte TRUSTSTORE_PATH und in der JDBC URL ändern. Im Folgenden finden Sie URL ein Beispiel für eine JDBC Verbindung für HS2 mit den erforderlichen Parametern:

```
jdbc:hive2://HOST_NAME:10000/
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD
```

Verwenden Sie unten die entsprechenden Anweisungen für die Verschlüsselung auf dem Cluster oder außerhalb von Cluster 2. HiveServer

On-cluster HS2 access

Wenn Sie mit dem Beeline-Client nach dem Primärknoten auf HiveServer 2 zugreifen, suchen Sie SSH anhand der Konfiguration `TRUSTSTORE_PATH` und `/etc/hadoop/conf/ssl-server.xml` nach den `TRUSTSTORE_PASSWORD` Parameterwerten.

```
ssl.server.truststore.location ssl.server.truststore.password
```

Die folgenden Beispielbefehle können Ihnen beim Abrufen dieser Konfigurationen helfen:

```
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

Wenn Sie von einem Client außerhalb des EMR Amazon-Clusters auf HiveServer 2 zugreifen, können Sie einen der folgenden Ansätze verwenden, um das und abzurufen: `TRUSTSTORE_PATH`
`TRUSTSTORE_PASSWORD`

- Konvertieren Sie die PEM Datei, die während der [Sicherheitskonfiguration](#) erstellt wurde, in eine JKS Datei und verwenden Sie dieselbe für die JDBC Verbindung. URL Verwenden Sie für openssl und keytool die folgenden Befehle:

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12 -destkeystore trustedCertificates.jks
```

- Alternativ können Sie auf `/etc/hadoop/conf/ssl-server.xml` verweisen, um die Parameterwerte `TRUSTSTORE_PATH` und `TRUSTSTORE_PASSWORD` mithilfe der Konfiguration `ssl.server.truststore.location` und `ssl.server.truststore.password` zu finden. Laden Sie die Truststore-Datei auf den Client-Computer herunter und verwenden Sie den Pfad auf dem Client-Computer als `TRUSTSTORE_PATH`.

Weitere Informationen zum Zugriff auf Anwendungen von einem Client außerhalb des EMR Amazon-Clusters finden Sie unter [Verwenden des JDBC Hive-Treibers](#).

Versionsverlauf von Hive

In der folgenden Tabelle sind die in jeder Release-Version von Amazon EMR enthaltenen Versionen von Hive sowie die mit der Anwendung installierten Komponenten aufgeführt. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Versionsinformationen von Hive

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-5.36,2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
		nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-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, 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-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, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-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, 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-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, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.3.2	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-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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httfs-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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httfs-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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httfs-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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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-httfs-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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Amazon EMR Release-Etikett	Hive-Version	Mit Hive installierte Komponenten
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

Hive-Versionshinweise nach Version

Themen

- [Amazon EMR 7.1.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.15.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.14.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.13.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.12.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.11.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.10.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.9.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.8.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.7.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.6.0 — Versionshinweise zu Hive](#)

Amazon EMR 7.1.0 — Versionshinweise zu Hive

Amazon EMR 7.1.0 — Hive-Änderungen

Typ	Beschreibung
Fehlerbehebung	HIVE-24381 — Die komprimierte Texteingabe gibt 0 Zeilen zurück, wenn die Kopf-/Fußzeile übersprungen wird.
Fehlerbehebung	HIVE-24190 —LLAP: ShuffleHandler könnte laut -4233 __ zurückgeben. DISK ERROR EXCEPTION TEZ
Fehlerbehebung	HIVE-23073 — Shade Netty .
Fehlerbehebung	HIVE-23073 — Shade Netty und Upgrade auf Netty 4.1.48.Final.

Typ	Beschreibung
Fehlerbehebung	HIVE-23148 — Der externe LLAP-Clientfluss ist aufgrund von Netty Shading unterbrochen.
Fehlerbehebung	HIVE-25180 — Aktualisiert Netty.
Fehlerbehebung	HIVE-24524 — LLAP ShuffleHandler: Führen Sie ein Upgrade auf Netty4 durch und entfernen Sie die Netty3-Abhängigkeit aus dem Hive, wo dies möglich ist.
Fehlerbehebung	HIVE-28000 — Hive QL: Die „not in“ - Klausel liefert falsche Ergebnisse, wenn kein Typenzwang stattfinden kann.
Fehlerbehebung	HIVE-27993 — Netty4 sollte einen Boss-Thread verwenden. ShuffleHandler
Upgrade	Führt ein Upgrade von Netty auf 4.1.100.Final durch
Upgrade	Führt ein Upgrade von Jetty auf 9.4.53.v20231009 durch
Upgrade	Führt ein Upgrade von Zookeeper auf 3.9.1 durch

Amazon EMR 7.1.0 — Hive-Änderungen

- Amazon EMR 7.1 aktualisiert Hive auf Netty 4.1.100.Final, um die Sicherheitslücken in Netty3 zu beheben. Da Hive von Netty3 abhängig ist, hive-druid-handler hat Hive den Klassenpfad von Hive hive-druid-handler JAR in Amazon 7.1 nicht. EMR Eine bevorstehende EMR Amazon-Version wird es in den Klassenpfad von Hive aufnehmen, sobald der Druid-Handler 4.1.100.Final oder spätere Versionen von Netty unterstützt. Wenden Sie sich an den AWS Support, wenn Sie das hive-druid-handler JAR in EMR Amazon-Versionen 7.1 oder höher benötigen.

Amazon EMR 6.15.0 — Versionshinweise zu Hive

Amazon EMR 6.15.0 — Hive-Änderungen

Typ	Beschreibung
Funktion	Support für TEZ-4397 — Für das asynchrone geteilte Öffnen von Tez unterstützt Hive jetzt die unter beschriebenen Tez-Konfigurationen. Asynchrone Öffnung von Tez-Splits
Fehlerbehebung	HIVE-25400 — Verschieben Sie die Offset-Aktualisierung nach. BytesColumnVector setValPreallocated
Fehlerbehebung	HIVE-25190 — Korrigiert viele kleine Zuordnungen in. BytesColumnVector
Upgrade	Apache Hadoop auf 3.3.6 aktualisieren.
Upgrade	HIVE-26684 — Upgrade auf 3.4.1. maven-shade-plugin
Verbesserung	Um die Startzeit des EMR Amazon-Clusters zu reduzieren, entfernen Sie 15 Sekunden Ruhezeit aus dem HCatalog Startskript.

Amazon EMR 6.14.0 — Versionshinweise zu Hive

Amazon EMR 6.14.0 — Hive-Änderungen

Typ	Beschreibung
Verbesserung	HIVE-26762 : Entfernen Sie das Kürzen von Operanden HiveFilterSetOpTransposeRule

Typ	Beschreibung
Fehlerbehebung	HIVE-27582 : Zwischenspeichern Sie das Tabelleneingabeformat nicht in HBase FetchOperator
Fehlerbehebung	HIVE-26452 : NPE bei der Konvertierung in eine Spalte, JOIN auf die mehr als einmal MAPJOIN verwiesen JOIN wird
Fehlerbehebung	HIVE-26416 : wirft nach Datei AcidUtils .isRawFormatFile() InvalidProtocolBufferException non-ORC
Fehlerbehebung	HIVE-26105 : „Spalten anzeigen“ zeigt zusätzliche Werte an, wenn die Spalte „Kommentare“ ein bestimmtes chinesisches Schriftzeichen enthält
Fehlerbehebung	HIVE-25864 : Die Hive-Abfrageoptimierung erzeugt einen falschen Plan für die Übertragung von Prädikaten mit Fensterfunktion
Fehlerbehebung	HIVE-25224 : Fehler bei mehreren Anweisungen mit Tabellen mit unterschiedlichen Ergebnissen INSERT bucketing_versions
Fehlerbehebung	HIVE-24151 : MultiDelimiterDe Verschiebt Daten, wenn Zeichenketten Zeichen enthalten non-ASCII
Fehlerbehebung	HIVE-23606 : () LLAP Verzögerung bei der Bereinigung für DirectByteBuffer EncodedReaderImpl

Typ	Beschreibung
Fehlerbehebung	HIVE-22165 : Die durch HIVE-14296 eingeführte Synchronisation verursacht eine hohe Latenz auf einem ausgelasteten Hive-Server <code>SessionManager.closeSession</code>
Fehlerbehebung	HIVE-21304 : Machen Sie die Nutzung der Bucketing-Version robuster

Amazon EMR 6.13.0 — Versionshinweise zu Hive

Amazon EMR 6.13.0 — Hive-Änderungen

Typ	Beschreibung
Verbesserung	Aktualisieren Sie Python-Skripte zur Support von Python3
Verbesserung	HIVE-27097 : Verbessern Sie die Wiederholungsstrategie für Client und Server MetaStore
Fehlerbehebung	HIVE-21778 CBO: „Struct is not null“ wird als Nullwert bewertet, was immer dazu führt, dass der Filter in der Abfrage fehlschlägt
Fehlerbehebung	HIVE-21009 : Dem Benutzer wurde die Möglichkeit hinzugefügt, einen Bind-Benutzer festzulegen
Fehlerbehebung	HIVE-22661 : Die Komprimierung schlägt bei einer Tabelle ohne Bucket fehl, deren Daten im Pfad geladen wurden
Fehlerbehebung	HIVE-19718 : Beim Hinzufügen von Partitionen in großen Mengen wird auch die Tabelle für jede Partition abgerufen

Typ	Beschreibung
Fehlerbehebung	HIVE-22173 : Eine Abfrage mit mehreren seitlichen Ansichten hängt während der Kompilierung
Fehlerbehebung	HIVE-27088 : Falsche Ergebnisse, wenn innere und äußere Joins mit Post-Join-Filtern zusammengeführt werden
Fehlerbehebung	HIVE-21935: Hive-Vektorisierung : Eingeschränkte Leistung beim Vektorisieren UDF
Fehlerbehebung	HIVE-25299 : Das Umwandeln von Zeitstempeln in numerische Datentypen ist für Nicht-Zeitzone falsch UTC
Fehlerbehebung	HIVE-24626 : LLAP: Leser-Threads könnten ausgehungert werden, wenn alle I/O-Elevator-Threads ausgelastet sind, um sich mit voller Warteschlange in die Warteschlange für andere Leser einzureihen
Fehlerbehebung	HIVE-27029 : Die Hive-Abfrage schlägt mit einem Fehler beim Schließen des Dateisystems fehl. Die Überarbeitung für -26352 wurde abgeschlossen HIVE
Fehlerbehebung	HIVE-26352 : Die Prüfung des Zugriffs auf die Tez-Warteschlange schlägt fehl und es wird eine Ausnahme bei der Komprimierung angezeigt GSS
Fehlerbehebung	HIVE-24590 : Bei der Protokollierung von Vorgängen gibt es immer noch Log4j-Appender
Fehlerbehebung	HIVE-24552 : Mögliche Verbindungen sind undicht oder häufen sich an HMS loadDynamicPartitions

Typ	Beschreibung
Fehlerbehebung	HIVE-27069 : Falsche Ergebnisse beim Bucket-Map-Join
Fehlerbehebung	HIVE-27344 : Fügt einen Nullcheck in #close hinzu RecordReaderImpl
Fehlerbehebung	HIVE-27439 : Support Leerzeichen in Dezimalzahlen
Fehlerbehebung	HIVE-27267 : Falsche Ergebnisse beim Bucket-Map-Join für eine Spalte mit Dezimalzahlen und Unterabfrage
Fehlerbehebung	HIVE-21986 : HiveServer Web-UI: Einstellung von Strict-Transport-Security im Standard-Antwort-Header
Fehlerbehebung	HIVE-22148 : S3A-Delegierungstoken werden nicht zur Job-Konfiguration des Compactors hinzugefügt.
Fehlerbehebung	HIVE-22622 : Hive ermöglicht das Erstellen einer Struktur mit doppelten Attributnamen
Fehlerbehebung	HIVE-22008 : LIKE Der Operator sollte einer mehrzeiligen Eingabe entsprechen
Fehlerbehebung	HIVE-23144 : Lass die Säuberung laufen LLAP QueryTracker serviceStop
Fehlerbehebung	HIVE-22391 : NPE Beim Überprüfen des Zwischenspeichers für Hive-Abfrageergebnisse
Fehlerbehebung	HIVE-23305 : NullPointerException aufgrund von Rennbedingungen drin LlapTaskSchedulerService addNode

Typ	Beschreibung
Fehlerbehebung	HIVE-22178 : Parkett wirft hinterher FilterPredicate CastException SchemaEvolution
Fehlerbehebung	HIVE-21517 : Reparieren AggregateStatsCache
Fehlerbehebung	HIVE-21825 : Verbessert die Client-Fehlermeldung, wenn Aktiv/Passive HA aktiviert ist
Fehlerbehebung	HIVE-23389 : kann dazu führen FilterMergeRule AssertionError
Fehlerbehebung	HIVE-22767 : Beeline analysiert Semikolons in Kommentaren nicht richtig
Fehlerbehebung	HIVE-22996 : BasicStats Beim Parsen sollte proaktiv nach Null oder einer leeren Zeichenfolge gesucht werden
Fehlerbehebung	HIVE-22808 : behandelt nicht HiveRelFieldTrimmer HiveTableFunctionScan
Fehlerbehebung	HIVE-22437 : LLAP Metadaten-Cache beim Sperren von Metadaten. NPE
Fehlerbehebung	HIVE-22606 : protokolliert avro.schema.literal unter Level AvroSerde INFO
Fehlerbehebung	HIVE-22713 : Für die Join-Fil (*) -RS-Struktur sollte keine konstante Weitergabe durchgeführt werden
Fehlerbehebung	HIVE-21624 : LLAP: Die CPU-Metriken auf Thread-Ebene sind defekt
Fehlerbehebung	HIVE-22815 : Reduziert die unnötige Erstellung von Dateisystemobjekten in MROutput

Typ	Beschreibung
Fehlerbehebung	HIVE-23060 : Die Abfrage schlägt mit dem Fehler fehl: „Der Ausdruck für Gruppierungssätze ist nicht im BY-Schlüssel enthalten. GROUP In der Nähe des Tokens ist ein Fehler aufgetreten
Fehlerbehebung	HIVE-22236 : Ansicht konnte nicht erstellt werden, wenn Ansicht ausgewählt wird, die Unterabfrage enthält NOT IN
Fehlerbehebung	HIVE-19886 : Logs können an zwei Dateien weitergeleitet werden, wenn —hiveconf hive.log.file verwendet wird
Fehlerbehebung	HIVE-20620 : Offensichtliche Kollisionen beim Einfügen in sortierte MM-Tabellen mit dynamischer Partitionierung
Fehlerbehebung	HIVE-14557 : Nullpointer, wenn sowohl als auch Mapjoin aktiviert sind SkewJoin
Fehlerbehebung	HIVE-20471 : Probleme beim Abrufen des Standard-Datenbankpfads
Fehlerbehebung	HIVE-20598 : Tippfehler in Berechnungen korrigieren HiveAlgorithmsUtil
Fehlerbehebung	HIVE-14737 : Problem beim Zugriff auf /logs in einer Kerberized Hive Server 2-Weboberfläche
Fehlerbehebung	HIVE-20733 : G NS darf = in Planbeschreibungen nicht verwenden enericUDFOPEqual
Fehlerbehebung	HIVE-20848 : Nach dem Einstellen schlägt die UpdateInputAccessTimeHook Abfrage mit der Meldung „Tabelle nicht gefunden“ fehl.

Typ	Beschreibung
Fehlerbehebung	HIVE-18929 : Die Methode <code>humanReadableInt</code> in <code>HiveStringUtils.java</code> hat eine Race-Bedingung.
Fehlerbehebung	HIVE-20841 : LLAP Dynamische Ports konfigurierbar machen
Fehlerbehebung	HIVE-20930 : <code>VectorCoalesce</code> Im FILTER Modus wird nicht wirksam
Fehlerbehebung	HIVE-21007 : Semi Join + Union kann zu falschen Plänen führen
Fehlerbehebung	HIVE-21074 : Das Bereinigen von Abfragen in Hive-Bucket-Tabellen funktioniert bei IS-Bedingungen nicht NOT NULL
Fehlerbehebung	HIVE-21223 : <code>CachedStore</code> gibt eine Null-Partition zurück, wenn die Partition nicht existiert
Fehlerbehebung	HIVE-19625 : Mögliche NPE und versteckte tatsächliche Ausnahme in <code>Hive# copyFiles</code>
Fehlerbehebung	HIVE-17020 : Aggressives RS-Deduplizieren kann fälschlicherweise den Zweig des OP-Baums entfernen
Fehlerbehebung	HIVE-20168 : Protokollierung ausgeblendet <code>ReduceSinkOperator</code>
Fehlerbehebung	HIVE-20879 : Die Verwendung von Null in einem Projektionsausdruck führt zu <code>CastException</code>
Fehlerbehebung	HIVE-20888 <code>TxnHandler</code> : <code>sort ()</code> wurde für unveränderliche Listen aufgerufen

Typ	Beschreibung
Fehlerbehebung	HIVE-19948 : teilt den Befehl nicht HiveCli richtig durch Semikolon auf, wenn die Zeichenfolge Anführungszeichen enthält
Fehlerbehebung	HIVE-20621 : Wird in resultset.next aufgerufen, was zu inkrementeller Langsamkeit führt GetOperationStatus
Fehlerbehebung	HIVE-20854 : Sinnvolle Standardeinstellungen: Das Zookeeper-Heartbeat-Intervall von Hive beträgt 20 Minuten, ändern Sie es auf 2
Fehlerbehebung	HIVE-20330 : HCatLoader Kann nicht mehrere Objekte für einen Job mit mehreren Eingaben verarbeiten InputJobInfo
Fehlerbehebung	HIVE-20787 : MapJoinBytesTableContainer dummyRow Case behandelt die Wiederverwendung nicht
Fehlerbehebung	HIVE-20331 : Die Abfrage mit Union All, Lateral View und Join schlägt fehl und gibt an, dass der Operator „Elternteil im untergeordneten Element nicht gefunden werden kann“
Fehlerbehebung	HIVE-19968 : Die UDF Ausnahme wird nicht gelöscht
Fehlerbehebung	HIVE-20410 : Abgebrochenes Einfügen und Überschreiben in der Transaktionstabelle verursacht den Fehler „Nicht genug Historie verfügbar für...“
Fehlerbehebung	HIVE-20059 : Hive-Streaming sollte in Ausnahmefällen bedingungslos das Shade-Präfix ausprobieren

Typ	Beschreibung
Fehlerbehebung	HIVENPE-19424 : Eingetragen MetaDataFormatters
Fehlerbehebung	HIVE-20355 : Parameter von bereinigen. HiveConnection setSchema
Fehlerbehebung	HIVE-20858 : Serializer wurde mit der Konfiguration in Utilities nicht korrekt initialisiert. createEmptyBuckets
Fehlerbehebung	HIVE-20424 : Das Schematool darf die Flugzeug-Historie nicht verunreinigen
Fehlerbehebung	HIVE-20338 : LLAP: Erzwingt synthetische Datei-IDs für Dateisysteme, die Protokoll-Impls mit Mutationssemantik haben HDFS POSIX
Fehlerbehebung	HIVE-11708 ClassCastExceptions : Logische Operatoren werden ausgelöst mit NULL
Fehlerbehebung	HIVE-21082 : InHPL/SQL unterstützt die Declare-Anweisung keine Variablen vom Typ Character
Fehlerbehebung	HIVE-16690 : Konfiguriert den kartesischen Tez-Produkttrand auf der Grundlage der Clustergröße LLAP
Fehlerbehebung	HIVE-21296: Beim Löschen der Varchar-Partition wird eine Ausnahme ausgelöst
Fehlerbehebung	HIVE OrInputFormat-14516 : SplitGenerator. callInternal
Fehlerbehebung	HIVE-20981 : Streaming/Lecks AbstractRecordWriter HeapMemoryMonitor

Typ	Beschreibung
Fehlerbehebung	HIVE-20043 : HiveServer 2: SessionState hat einen statischen Sync-Block um einen AtomicBoolean
Fehlerbehebung	HIVE-20191 : Die PreCommit Patch-Anwendung schlägt nicht fehl, wenn der Patch leer ist
Fehlerbehebung	HIVE-20400 : Die Tabelle erstellen sollte immer einen vollqualifizierten Pfad verwenden, um mögliche FS-Mehrdeutigkeiten zu vermeiden
Fehlerbehebung	Fügen Sie vor dem Zugriff auf schiefe Spalten eine Nullprüfung für skewedInfo hinzu

Amazon EMR 6.12.0 — Versionshinweise zu Hive

Amazon EMR 6.12.0 — Hive-Änderungen

Typ	Beschreibung
Verbesserung	Support für JDK 11 und JDK 17 Runtime hinzugefügt
Verbesserung	Unterstützung für die Abfrage von Spaltennamen mit Berücksichtigung von Groß- und Kleinschreibung und reservierten Schlüsselwörtern hinzugefügt, wenn S3 Select verwendet wird. Um es zu verwenden, definieren Sie die Tabelleneigenschaft im Format „s3select .column.mapping“ = " <i>column1:fieldName1</i> , <i>column2:fieldName2</i> , ..."
Verbesserung	HIVE-23133 : Numerische Operationen können je nach Hardware zu unterschiedlichen Ergebnissen führen

Typ	Beschreibung
Verbesserung	HIVE-27145 : Wird StrictMath für die übrigen mathematischen Funktionen als Nachfolger von -23133 verwendet HIVE
Fehlerbehebung	Die Wildcard-Inkompatibilität in get_partitions_by_filter und get_num_partitions_by_filter wurde behoben, die durch die Portierung von -22900 in Hive 6.4.0 verursacht wurde HMS APIs HIVE EMR
Fehlerbehebung	HIVE-26736 : Autorisierungsfehler für verschachtelte Views mit Klausel WITH
Fehlerbehebung	HIVE-22416 : MR-bezogene Operationsprotokolle fehlen, wenn die parallel Ausführung aktiviert ist
Fehlerbehebung	HIVE-19653 : Falscher Prädikat-Pushdown für Groupby mit Gruppierungssätzen
Fehlerbehebung	HIVE-22094 : Abfragen schlagen fehl mit: hive.ql.exec.vector. ClassCastException DecimalColumnVectorkann nicht in Hive.QL.Exec.Vector.Decimal64 umgewandelt werden ColumnVector
Fehlerbehebung	HIVEPTF-26340 : Der vektorisierte Operator schlägt fehl, wenn die Abfrage eine Fensterfunktion in Großbuchstaben hat
Fehlerbehebung	HIVE-26184 : COLLECT _ SET mit GROUP BY ist sehr langsam, wenn einige Tasten stark schief sind
Fehlerbehebung	HIVE-26373 : ClassCastException beim Lesen von Zeitstempeln aus einer Tabelle mit Avro-Daten HBase

Typ	Beschreibung
Fehlerbehebung	HIVE-26388 : ClassCastException wenn es in der Quelltable der CTAS Abfrage eine Spalte gibt, die keine Zeichenfolge ist. Upgrade HIVE-26172: Hive — Upgrade von Ant auf 1.10.11 aufgrund von -2021-36373 und -2021-36374 CVE CVE
Fehlerbehebung	HIVE-26114 : Wenn die JDBC-Verbindung hiveserver2 mit dem Befehl dfs mit Präfixraum repariert wird, wird dies zu einer Ausnahme führen
Fehlerbehebung	HIVE-26396 : Die Trunc-Funktion hat ein Problem mit dem Präzisionsabfangen und das Ergebnis hat viele 0
Fehlerbehebung	HIVE-26446 : Das Feld für partitionierte Tabellen HiveProtoLoggingHook kann nicht gefüllt werden. TablesWritten
Fehlerbehebung	HIVE-26639 : und sollte sich nicht auf den Standardzeichensatz verlassen ConstantVectorExpression ExplainTask
Fehlerbehebung	HIVE-22670 : ArrayIndexOutOfBoundsException wenn der vektorisierte Reader zum Lesen einer Parquet-Datei verwendet wird
Fehlerbehebung	HIVE-23607 : Berechtigungsproblem: Das Erstellen einer Ansicht in einer anderen Ansicht ist erfolgreich, aber das Ändern der Ansicht schlägt fehl
Fehlerbehebung	HIVE-25498 : Eine Abfrage mit mehr als 31 verschiedenen Funktionen gibt ein falsches Ergebnis zurück

Typ	Beschreibung
Fehlerbehebung	HIVE-25780 : DistinctExpansion erstellt mehr als 64 Gruppierungssätze II
Fehlerbehebung	HIVE-23868 : Spezifikation der Fensterfunktion: unterstützt 0 vorne/nach
Fehlerbehebung	HIVE-24539 : Die Schemagenerierung sollte das Spaltentrennzeichen berücksichtigen OrclInputFormat
Fehlerbehebung	HIVE-23476 : LLAP Ordnet Arenen auch für die Groß-/Kleinschreibung der Karten vorab zu
Fehlerbehebung	HIVE-25806 : Mögliches Leck in - Parquet, IO LlapCacheAwareFs LLAP
Fehlerbehebung	HIVE-23498 : Trace-Methode deaktivieren bei HTTP ThriftHttpClientService
Fehlerbehebung	HIVE-25729 : ThriftUnionObjectInspector sollte benachrichtigt werden, wenn es vollständig initiiert ist
Fehlerbehebung	HIVE-23846 : Vermeiden Sie unnötige Serialisierung und Deserialisierung von Bitvektoren
Fehlerbehebung	HIVE-24233 : außer dass die Unterabfrage bei deaktiviertem CBO einen Nullzeiger auslöst
Fehlerbehebung	HIVE-24276 : Sicherheitslücke in loggerconf.jsp Cross-Site Scripting () HiveServer XSS
Fehlerbehebung	HIVE-25721 : Das Ergebnis des äußeren Joins ist falsch

Typ	Beschreibung
Fehlerbehebung	HIVE-25223 : Auswahl mit Limit gibt keine Zeilen in einer nicht systemeigenen Tabelle zurück
Fehlerbehebung	HIVE-25794 CombineHiveRecordReader: Log-Anweisungen in einer Schleife führen zu Speicherdruck
Fehlerbehebung	HIVE-23602 : Verwenden Sie das Java Concurrent Package für Operation Handle Set
Fehlerbehebung	HIVE-24045 : Keine Protokollierung im Zusammenhang mit dem Zeitpunkt, zu dem die Standarddatenbank erstellt wurde
Fehlerbehebung	HIVE-24305 : Das Avro-Dezimalschema füllt Skala/Genauigkeit nicht richtig aus, wenn der Wert in Anführungszeichen eingeschlossen ist
Fehlerbehebung	HIVE-25844 : Deserialisierungsfehler in Ausnahmefällen können dazu führen, dass Beeline sofort beendet wird
Fehlerbehebung	HIVE-25040: Die Drop-Datenbank-Kaskade kann persistente Funktionen nicht entfernen
Fehlerbehebung	HIVE-23501 : VectorDeserializeRow wenn komplexe Typen AOOB in primitive Typen umgewandelt werden
Fehlerbehebung	HIVE-23704 : Thrift HTTP Server verarbeitet das Auth-Handle nicht korrekt
Fehlerbehebung	HIVE-23529 : ist für Uniontype kaputt, wenn row_deserialize CTAS

Typ	Beschreibung
Fehlerbehebung	HIVE-24144 : getIdentifierQuote Die HiveDatabaseMetaData Zeichenfolge in gibt einen falschen Wert zurück
Fehlerbehebung	HIVE-23850 : Erlaubt, PPD wenn der Betreff keine Spalte mit vorhandenen Gruppierungssätzen ist
Fehlerbehebung	HIVE-24036 : Kryo-Ausnahme beim Serialisieren des Anrufplans getSplits UDF
Fehlerbehebung	HIVE-25919 : beim Eingeben eines booleschen Spaltenprädikats ClassCastException HBaseStorageHandler
Fehlerbehebung	HIVE-25261 : RetryingHMSHandler sollte das MetaException mit einer kurzen Beschreibung des Ziels umschließen
Fehlerbehebung	HIVE-24792 : Potenzielles Thread-Leck während des Betriebs
Fehlerbehebung	HIVE-23409 : Wenn das erneute Öffnen der TezSession Anwendung fehlschlägt, weil der Timeline-Dienst ausgefallen ist, wird das Standardformular nach TezSession einem SessionPool erneuten Versuch geschlossen
Fehlerbehebung	HIVE-23615 : Verweisen Sie nicht auf Nullzeiger in der Beeline-Befehlsklasse
Fehlerbehebung	HIVE-24849 : Erzeugt ein Socket-Timeout für externe Tabellen, wenn der Speicherort eine große Anzahl von Dateien enthält (betrifft 3.1.2)

Typ	Beschreibung
Fehlerbehebung	HIVE-24193 : Die Auswahlabfrage für die umbenannte Hive-Acid-Tabelle erzeugt keine Ausgabe
Fehlerbehebung	HIVE-25209 : SELECT Abfrage mit Funktion, die ein unerwartetes Ergebnis erzeugt SUM
Fehlerbehebung	HIVE-23666 : checkHashMode Effizienz wird übersprungen, wenn für einen Groupby-Operator kein Gruppierungssatz festgelegt ist
Fehlerbehebung	HIVE-23873 : Die Abfrage der Hive-Tabelle schlägt fehl, wenn sie ausgeschaltet ist JDBCStorageHandler NPE CBO
Fehlerbehebung	HIVE-24149 : Schließt die Verbindung nicht HiveStreamingConnection HMS
Fehlerbehebung	HIVE-25561 : Die getötete Aufgabe sollte die Datei nicht übertragen. (betrifft die Versionen 2.x und 3.x)
Fehlerbehebung	HIVE-25683 : Schließt den Leser. AcidUtils isRawFormatDatei
Fehlerbehebung	HIVE-24294 : TezSessionPool Sessions können ausgelöst werden AssertionError
Fehlerbehebung	HIVE-24182 : Ranger-Autorisierungsproblem mit permanentem Status UDFs
Fehlerbehebung	HIVE-22805 : Die Vektorisierung mit bedingtem Array oder Map ist nicht implementiert und löst einen Fehler aus

Typ	Beschreibung
Fehlerbehebung	HIVE-22828: Decimal64: und Anweisungen konvertieren implizit decimal64 in 128 NVL CASE
Fehlerbehebung	HIVE-21398 : Spalten mit geschätzten Statistik en sollten nicht als eindeutige Schlüssel betrachtet werden
Fehlerbehebung	HIVE-22490 : Das Hinzufügen von Jars mit Sonderzeichen in ihrem Pfad führt zu einem Fehler
Fehlerbehebung	HIVE-22700 : Bei nicht autorisierten Komprimie rungen kann Speicherplatz verloren gehen
Fehlerbehebung	HIVE-22053 : Der Funktionsname wird beim Erstellen der Funktion nicht normalisiert
Fehlerbehebung	HIVE-22595 : Dynamische Partitionseinfügun gen schlagen in einer Avro-Tabellentabelle mit externem Schema fehl
Fehlerbehebung	HIVE-21795 : Die Zeile mit der Rollup-Zu sammenfassung fehlt möglicherweise, wenn ein Mapjoin in einer partitionierten Tabelle stattfind et
Fehlerbehebung	HIVE-22987: In wann ist Null ClassCast Exception VectorCoalesce DataTypeP hysicalVariation
Fehlerbehebung	HIVE-22814 : bei der Vektorisierung ArrayInde xOutOfBound getDataType PhysicalVariation
Fehlerbehebung	HIVE-22523 : Der Fehlerhandler in blockiert LlapRecordReader möglicherweise, wenn seine Warteschlange voll ist

Typ	Beschreibung
Fehlerbehebung	HIVE-21796 : ArrayWritableObjectInspector.equals kann O (2^nesting_depth) Zeit in Anspruch nehmen
Fehlerbehebung	HIVE-22929 : Leistung: Beim Parsen von Bezeichnern in Anführungszeichen wird Regexp per String verwendet. replaceAll()
Fehlerbehebung	HIVE-21641 : Der externe Llap-Client gibt Dezimalspalten mit unterschiedlicher Genauigkeit/Skalierung zurück als Beeline
Fehlerbehebung	HIVE-22207 : Tez: wird ausgelöst, wenn „dfs.block.size“ auf dem Cluster „128m“ SplitGenerator ist NumberFormatException
Fehlerbehebung	HIVE-22114 : Das Einfügen einer Abfrage für eine partitionierte Tabelle, die nur zum Einfügen einfügt, schlägt fehl, wenn alle Buckets leer sind
Fehlerbehebung	HIVE-22165 : Die Synchronisation wurde mit -14296 eingeführt. HIVE SessionManager closeSession verursacht eine hohe Latenz auf einem ausgelasteten Hive-Server
Fehlerbehebung	HIVE-22744 : TezTask Für den Scheitelpunkt mit mehr als einem Rand sollte ein proportionaler Sortierspeicher vorhanden sein
Fehlerbehebung	HIVE-22072 : Wenn Sie die Tabelle ändern, um eine Spalte zu ändern, werden die Beschränkungsreferenzen nicht aktualisiert
Fehlerbehebung	HIVE-22075 : Korrigiert die Regression max-reducers=1 von -14200 HIVE

Typ	Beschreibung
Fehlerbehebung	HIVE-22527 : Hive on Tez: Der Job zum Zusammenführen kleiner Dateien wird in eine andere Warteschlange gestellt (Standard-Warteschlange)
Fehlerbehebung	HIVE-22816 : QueryCache: Abfragen, die Views verwenden, können nach der Erweiterung zwischengespeichert werden CTE
Fehlerbehebung	HIVE-22733 : Nach dem Deaktivieren der Betriebsprotokolleigenschaft in Hive wird das Betriebsprotokoll trotzdem gespeichert HS2
Fehlerbehebung	HIVE-22699 : Die Maske sollte den numerischen Wert 0 maskieren UDFs
Fehlerbehebung	HIVE-23356 : Die Hash-Aggregation ist bei der Verarbeitung von Abfragen mit Gruppierungsausdrücken immer deaktiviert.
Fehlerbehebung	HIVE HiveRelOptUtil-21568 : isRowFilteringDer Plan sollte das Projekt überspringen
Fehlerbehebung	HIVE-21760 : Die Optimierung für gemeinsame Arbeit sollte bei Joins umgangen werden SMB
Fehlerbehebung	HIVE-22712 : ReExec Führt den Treiber aus und sendet die Abfrage unabhängig von der benutzerdefinierten Warteschlange in der Standardwarteschlange
Fehlerbehebung	HIVE-21397 : BloomFilter Für Hive funktioniert die Tabelle Managed [ACID] nicht wie erwartet
Fehlerbehebung	HIVE-23011 : Der Shared Work Optimizer sollte beim Vergleich von Joins die restlichen Prädikate überprüfen

Typ	Beschreibung
Fehlerbehebung	HIVE-21412 : funktioniert nicht mit Tabellen PostExecOrcFileDump ACID
Fehlerbehebung	HIVE-22201 : ConvertJoinMapJoin # wird ausgelöst, ArrayIndexOutOfBoundsException wenn checkShuffleSize ForLargeTable keine große Tabelle ausgewählt ist
Fehlerbehebung	HIVE-21971 : Der Classloader lässt aufgrund von `:ReflectionUtils: CONSTRUCTOR _CACHE` mit temporären Funktionen HS2 durchsickern + Generic UDF
Fehlerbehebung	HIVE-21368 : Vektorisierung: Unnötiges Decimal64 -> Konvertierung HiveDecimal
Fehlerbehebung	HIVE-25416: Hive-Metastore-Speicherleck aufgrund eines Fehlers datanucleus-api-jdo
Fehlerbehebung	HIVE-22219: Das Herunterfahren eines Node-Managers blockiert den Neustart des Dienstes LLAP
Fehlerbehebung	HIVE-21793 : CBO Ruft Spaltenstatistiken ab, auch wenn hive.stats.fetch.column.stats auf false gesetzt ist
Fehlerbehebung	HIVE-22163 :: Durch die Aktivierung wird die Statistikschatzung aktiviert, auch wenn die Schätzung deaktiviert ist CBO CBO
Fehlerbehebung	HIVE-18735 : Beim Erstellen einer Tabelle wird das Transaktionsattribut verloren

Typ	Beschreibung
Fehlerbehebung	HIVE-22433 : Hive JDBC Storage Handler: Falsche Ergebnisse wurden von und aus der Datenquelle abgerufen BOOLEAN TIMESTAMP DataType JDBC
Fehlerbehebung	HIVE ObjectStore-19430 : cleanNotificationEvents OutOfMemory bei einer großen Anzahl von ausstehenden Veranstaltungen
Fehlerbehebung	HIVE-20785 : Falscher Schlüsselname in. JDBC DatabaseMetaData getPrimaryKeys Methode
Fehlerbehebung	HIVE-16116 : Beeline wird ausgelöst, NPE wenn beeline.hiveconfvariables= {} in beeline.properties
Fehlerbehebung	HIVE-20066 : hive.load.data.owner wird mit dem vollständigen Prinzipal verglichen
Fehlerbehebung	HIVE-20489 : Erläutern Sie, dass der Abfrageplan nicht mehr funktioniert
Fehlerbehebung	HIVE-21033 : Wenn vergessen wird, den Vorgang zu beenden, werden weitere 2 Ausgaben unterbrochen HiveServer
Fehlerbehebung	HIVE-19888 : Irreführende Warnung „METASTORE_ FILTER _ HOOK wird ignoriert“ von SessionState
Fehlerbehebung	HIVE-20303 : db.table (...) INSERT OVERWRITE TABLE PARTITION NOT EXISTS IF wirft InvalidTableException
Fehlerbehebung	HIVE-16144 : CompactionInfo hat kein Äquivalens, wird aber im Set hashCode verwendet

Typ	Beschreibung
Fehlerbehebung	HIVE-20818 : Mit einer WHERE Unterabfrage erstellte Ansichten betrachten Ansichten, auf die in der Unterabfrage verwiesen wird, als direkte Eingabe
Fehlerbehebung	HIVE-21005 : Mehr Stripes pro Split-Leaks lesen LLAP ZlibCodecs
Fehlerbehebung	HIVE-20771 : schlägt bei leeren Strukturen fehl. LazyBinarySerDe
Fehlerbehebung	HIVE-18852 : Irreführende Fehlermeldung bei der Validierung von Änderungen an Tabellen
Fehlerbehebung	HIVE-21124 :HPL/SQLunterstützt die Anweisung nicht CREATE TABLE LIKE
Fehlerbehebung	HIVE-20935 : Der Upload des Llap-Paket-Tarballs schlägt fehl und verursacht einen Fehler beim Starten des Dienstes EC2 LLAP
Fehlerbehebung	HIVE-20409 : HiveACID: Durch Aktualisieren/Löschen/Zusammenführen wird das HDFS-Staging-Verzeichnis nicht bereinigt
Fehlerbehebung	HIVE-20570 : Die Vereinigung mit ALL hive.optimize.union.remove=true hat einen falschen Plan
Fehlerbehebung	HIVE-20421 : Unzulässige Zeichen-Entität '\ b' in hive-default.xml.template
Fehlerbehebung	HIVE-19133 : Die phasenweisen Leistungskennzahlen der HS2 WebUI werden nicht korrekt angezeigt

Typ	Beschreibung
Fehlerbehebung	HIVE-18977 : Beim Auflisten von Partitionen werden unterschiedliche Ergebnisse mit und Direct zurückgegeben JDO SQL
Fehlerbehebung	HIVE-20034 : Aus Gründen der Abwärtskompatibilität werden Änderungen bei der MetaStore Ausnahmebehandlung rückgängig gemacht
Fehlerbehebung	HIVE-20672 : Der Thread, der sich anmeldet, sollte jedes feste Intervall melden LlapTaskSchedulerService
Fehlerbehebung	HIVE-12812 : Aktiviert mapred.input.dir.recursive standardmäßig, um Union mit Aggregatfunktion zu unterstützen
Fehlerbehebung	HIVE-20147: Hive-Streaming-Ingest begnügt sich mit synchronisierter Protokollierung
Fehlerbehebung	HIVE-19203: Thread-Sicherheitsproblem in HiveMetaStore
Fehlerbehebung	HIVE-20091 : Tez: Sicherheitsanmeldedaten für die Ausgabe hinzufügen FileSinkOperator
Fehlerbehebung	HIVE-16906 : Hive ATSHook sollte nach yarn.timeline-service.enabled suchen, bevor eine Verbindung hergestellt wird ATS
Fehlerbehebung	HIVE-20714 : tblproperties für eine einzelne Eigenschaft gibt den Wert in der Namensspalte zurück SHOW
Fehlerbehebung	HIVE-24730 : Shims-Klassen überschreiben Werte aus hive-site.xml und tez-site.xml im Hintergrund

Typ	Beschreibung
Fehlerbehebung	HIVE-22055 : Die gewählte Anzahl liefert nach dem Laden von Daten aus einer Textdatei ein falsches Ergebnis

Amazon EMR 6.11.0 — Versionshinweise zu Hive

Amazon EMR 6.11.0 — Hive-Änderungen

Typ	Beschreibung
Verbesserung	Es wurde Unterstützung für das Löschen von Partitionen mit mehreren Threads hinzugefügt, um die Leistung beim Löschen von Partitionen zu verbessern
Verbesserung	Support das Lesen von codierten Hive-Abfrage-dateien
Verbesserung	Der Tez Shuffle Handler ist standardmäßig für Hive-on-Tez-Jobs aktiviert
Fehler	Es wurde eine Option hinzugefügt, um die deterministische Verteilung von Schlüsseln an Reducer zu aktivieren, um falsche Ergebnisse zu korrigieren, wenn hive.groupby.skewindata aktiviert ist (gemeldet in -20220) HIVE
Fehler	Fehler bei der Statistikberechnung behoben, wenn der Standardpartitionsname konfiguriert war
Fehler	Beachten Sie alle benutzerdefinierten SSL Klassifizierungsparameter, die bei der standardmäßigen Konfiguration für 2 in einem Cluster mit SSL aktivierter Verschlüsselung

Typ	Beschreibung
	während der Übertragung übergeben wurden HiveServer
Backport	HIVE-23617 : Probleme mit der Speicher-API wurden behoben FindBug
Backport	HIVE-26408 : Vektorisierung: Die Deallokation von Scratch-Spalten wurde korrigiert, untergeordnete Spalten werden nicht als Ausgabe wiederverwendet ConstantVectorExpression
Backport	HIVE-23614 : Immer weiterleiten an HiveConfig removeTempOr DuplicateFiles
Backport	HIVE-23354 : Entferne die Überprüfung der Dateigrößensicherheit von compareTempOr DuplicateFiles
Backport	HIVE-20344 : Fehler beim Werfen behoben . PrivilegeSynchronizer SBA AccessControlException Außerdem wurde die Eigenschaft hive.privilege.synchronizer eingeführt, um den Privilege Synchronizer zu deaktivieren
Backport	HIVE-15826 : Support die Konfiguration von 'serialization.encoding' für alle SerDes
Backport	HIVE-18284 : Fehler NPE beim Einfügen von Daten mit der 'Distribute-By'-Klausel mit Dynpart-Sortieroptimierung behoben
Backport	HIVE-24930 : Operator. setDone() Der Kurzschluss von der untergeordneten Operation wird im vektorisierten Codepfad nicht verwendet (wenn == 1) childSize

Typ	Beschreibung
Backport	HIVE-24523 : Der vektorisierte Lesepfad für berücksichtigt nicht den For-Zeitstempel LazySimpleSerde SERDEPROPERTIES
Backport	HIVE-23265 : Doppelte Rowsets werden zurückgegeben, wenn Limit und Offset gesetzt sind
Backport	HIVE-21492 : Die mit dem Thrift/Custom-Tool generierte Parquet-Datei VectorizedParquetRecordReader kann nicht gelesen werden
Backport	HIVE-22540: Vektorisierung: Decimal64-Spalten funktionieren nicht mit. Vectorize dBatchUtil makeLikeColumnVektor ()
Backport	HIVE-22588: Leert die verbleibenden Zeilen für die restlichen Gruppierungssätze, wenn der Gruppierungsmodus des Vektors umgeschaltet wird
Backport	HIVE-22551 : sollte Vektor und Länge konsistent bereinigen BytesColumnVector initBuffer
Backport	HIVE-22448 :CBO: Erweitert die eindeutige Mehrfachzählung mit einem Gruppierungsschlüssel
Backport	HIVE-22248 : Behebt anhaltende Probleme mit Statistiken
Backport	HIVE-22210 : Bei der Vektorisierung können beim Filtern verwendete Berechnungsausgabespalten wiederverwendet werden
Backport	HIVE-21531: Vektorisierung: Alle Hashcodes werden nicht mit Murmur3 berechnet NULL

Typ	Beschreibung
Backport	HIVE-20419 : Vektorisierung: Verhindert die Mutation von nach der Verwendung in einem Hashmap-Schlüssel VectorPartitionDesc
Backport	HIVE ClassCastException -19388 VectorMap JoinCommonOperator : während der Initialisierung
Backport	HIVE-21584 : Java 11-Vorbereitung: Der Systemklassenlader ist nicht URLClassLoader
Backport	HIVE-25107 : Die Klassenpfad-Protokollierung sollte sich auf Level befinden (#2271) DEBUG
Backport	HIVE-22097 : Inkompatibles java.util. ArrayList für Java 11
Backport	HIVE-23938 : LLAP: JDK11 - Einige JVM-Argumente, die sich auf die Rotation von GC-Logdateien beziehen, können nicht mehr verwendet werden
Backport	HIVE-26226 : Schließt jdk.tools dep aus Hive-Metastore in Upgrade-Acid aus
Backport	HIVE-17879 : Aktualisiere das Datanucleus Maven-Plugin
Backport	HIVE-27004 : DateTimeFormatterBuilder # appendZoneText kann '+' in Java-Versionen höher als 8 nicht parsen UTC
Backport	HIVE-16812 : filtert keine gelöschten Ereignisse VectorizedOrcAcidRowBatchReader
Backport	HIVE-17917 : VectorizedOrcAcidRowBatchReader computeOffsetAndBucket-Optimierung

Typ	Beschreibung
Backport	HIVE-19985 ACID: Überspringe die Dekodierung der ROW __ID-Abschnitte für schreibgeschützte Abfragen
Backport	HIVE-20635 : VectorizedOrcAcidRowBatchReader filtert keine Löschereignisse für Originaldateien
Upgrade	Upgrade auf Javadoc 3.3.1
Upgrade	Aktualisieren Sie Javassist auf 3.24.1-GA
Upgrade	Update auf 2.0.0-M14 apache-directory-server

Neue Konfigurationen

Name	Klassifizierung	Beschreibung
hive.metastore.fs.drop.partition.threads	hive-site	Anzahl der Core-Threads im Threadpool der Drop-Partition.
hive.metastore.fs.drop.partition.keepalive.time	hive-site	Zeit in Sekunden, in der ein asynchroner Drop-Partition-Thread (aus dem Thread-Pool) im Leerlauf auf das Eintreffen einer neuen Aufgabe wartet, bevor er beendet wird.
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	Maximale Warteschlangenlänge, die im Thread-Pool zum Löschen von Partitionen aus dem Dateisystem verwendet werden soll.
hive.groupby.enable.deterministic.distribution	hive-site	Ermöglichen Sie die deterministische Verteilung von

Name	Klassifizierung	Beschreibung
		Schlüsseln an Reducer. Beim Aufrufen der Funktion Rand, die für die zufällige Partitionierung verwendet wird, wird ein konstanter Startwert übergeben.
hive.privilege.synchronizer	hive-site	Ob die Rechte eines externen Autorisierers in 2 regelmäßig synchronisiert werden sollen. HiveServer
hive.cli.query.file.encoding	hive-site	Dateikodierung für alle Arten von Abfragedateien (Abfragedatei, Init-Abfragedatei, RC-Datei usw.), die in den CLI-Argumenten bereitgestellt werden.
hive.emr.tez.shuffle.enabled	hive-site	Hive-on-Tez-Aufträge verwenden jetzt standardmäßig tez_shuffle anstelle von mapreduce_shuffle als Standard-Shuffle-Handler.

Veraltete Konfigurationen

Die folgenden Konfigurationseigenschaften sind aufgrund von [HIVE-23354](#) veraltet und werden in EMR Amazon-Versionen 6.11.0 und höher nicht mehr unterstützt.

Name	Standardwert
hive.mapred.reduce.tasks speculative.execution	false
tez.am.speculation.enabled	false

Amazon EMR 6.10.0 — Versionshinweise zu Hive

Amazon EMR 6.10.0 — Hive-Änderungen

Typ	Beschreibung
Funktion	Aktivieren Sie die AWS Lake Formation basierte Zugriffskontrolle für Apache Hive-Abfragen (Schreiben) über IAM Passthrough (Hive / Steps). CLI API
Verbesserung	Deaktivieren Sie die Konfiguration <code>hive.log.explain.output</code> standardmäßig, um die Protokollgröße zu reduzieren
Backport	HIVE-26408 : Vektorisierung: Die Deallokation von Scratch-Spalten wurde korrigiert, untergeordnete Spalten werden nicht als Ausgabe wiederverwendet ConstantVectorExpression
Backport	HIVE-22269 : Die falsche Anzahl von Reduktoren in Einfügeabfragen mit dynamischer Partition wurde behoben, weil Statistiken aufgrund von -20703 fehlten. HIVE
Backport	HIVE-22891 : Extrahierung bei Nichtausführung überspringen PartitionDesc CombineHiveRecord LLAP
Backport	HIVE-23804 : Fügt eine Standarddatenbank für spaltenstatistikspezifische Tabellen im Hive-Metastore-Schema hinzu, um sie abwärtskompatibel zu machen
Backport	HIVE-25277 : Langsames Löschen von Hive-Partitionen für Cloud-Objektspeicher mit teuren Kosten ListFiles

Typ	Beschreibung
Backport	HIVE-19202 : CBO Fehler aufgrund von in. NullPointerException HiveAggregate isBucketedInput()
Backport	HIVE-19048 : Beeline beheben Initscript-Fehler werden ignoriert
Backport	HIVE-21085 : Die Registrierung für Materialized Views startet eine nicht externe TEZ-Sitzung
Backport	HIVE-21675 : CREATE VIEW IF NOT EXISTS gibt statt „OK“ einen Fehler zurück, wenn die Ansicht bereits existiert. Dies ist eine Regression von Hive 2.
Backport	HIVE-21646 : Tez: Verhindert TezTasks, dass dem Kontext der Thread-Protokollierung entgangen wird
Backport	HIVE-22054 : Vermeidet rekursives Auflisten, um zu überprüfen, ob ein Verzeichnis leer ist
Backport	HIVE-16587 : NPE beim Einfügen komplexer Typen mit verschachtelten Nullwerten
Backport	HIVE-22647 : Aktiviert standardmäßig den Sitzungspool
Backport	HIVE-13288 : Verwirrende Ausnahmemeldung in. DagUtils localizeResource
Backport	HIVE-23870 : Optimieren Sie mehrere Textkonvertierungen in. WritableHiveCharObjectInspector getPrimitiveJavaObjekt/HiveCharWritable
Backport	HIVE-21498 : Aktualisiere Thrift auf 0.13.0

Typ	Beschreibung
Backport	HIVE-24378 : Leerzeichen am Anfang und am Ende werden vor der Dezimalkonvertierung nicht entfernt
Backport	HIVE-21341 : Sinnvolle Standardwerte: hive.server2.idle.operation.timeout und hive.server2.idle.session.timeout sind zu hoch
Backport	HIVE-22465 : Fügen Sie die SSL-Konfiguration hinzu TezConfigurationFactory
Backport	HIVE-24710 : Optimieren Sie die PTF Iteration für count (*), um die I/O-Kosten zu reduzieren CPU
Backport	HIVE-15406 : Erwägen Sie die Vektorisierung der neuen Funktion „Trunc“
Backport	HIVE-21541 : Korrigiert fehlende ASF-Header von -15406 HIVE
Backport	HIVE-24808 : Analysierte Daten zwischenspeichern
Backport	HIVE-24746 : PTF: TimestampValueBoundaryScanner kann während der Bereichsberechnung optimiert werden
Backport	HIVE-25059 : Das Alter-Ereignis wird während der Replikation in ein Umbenennungereignis umgewandelt
Backport	HIVE-25142 : Beim erneuten Hash in der Map-Join-Fast-Hash-Tabelle werden große Schlüssel beschädigt

Typ	Beschreibung
Backport	HIVE-23756 : Der package.jdo-Datei wurden weitere Einschränkungen hinzugefügt
Backport	HIVE-25150 : Tabulatorzeichen werden vor der Dezimalkonvertierung nicht entfernt, ähnlich wie Leerzeichen, das als Teil von -24378 korrigiert wurde HIVE
Backport	HIVE-25093 : date_format () gibt nur die Ausgabe in der Zeitzone zurück UDF UTC
Backport	HIVE-25268 : date_format udf gibt falsche Ergebnisse für Daten vor 1900 zurück, wenn die lokale Zeitzone eine andere ist als UTC
Backport	HIVE-25338 : in conv, wenn die Eingabe leer ist AIOBE UDF
Backport	HIVE-22400 : Die Minute mit der Zeit kehrt zurück UDF NULL
Backport	HIVE-25058 :PTF: TimestampValueBoundaryScanner kann bei der Reichweitenberechnung optimiert werden pt2 - isDistanceGreater
Backport	HIVE-25449 : datediff () gibt eine falsche Ausgabe aus, wenn es in einer TEZ-Aufgabe mit einer anderen Zeitzone ausgeführt wird UTC
Backport	HIVE-23688 : Vektorisierung: Für eine Spalte vom Typ Map, die einen Nullwert enthält IndexArrayOutOfBoundsException
Backport	HIVE-22247 : HiveHFile OutputFormat wird ausgelöst, wenn die Task-Ausgabe der Partition leer ist FileNotFoundException

Typ	Beschreibung
Backport	HIVE-25570 : Hive sollte den vollständigen URL Pfad zur Autorisierung für den Befehl insert overwrite location senden
Backport	HIVE-22903 : Die vektorisierte row_number () setzt die Zeilennummer nach einem Batch zurück, falls ein konstanter Ausdruck in der Partitionsklausel vorhanden ist
Backport	HIVE-25549 : Falsche Ergebnisse für eine Fensterfunktion mit einem Ausdruck in der BY- oder BY-Klausel PARTITION ORDER
Backport	HIVE-25579 : LOAD Überschreiben fügt statt Überschreiben
Backport	HIVE-25659 : Metastore-Direkt-SQL-Abfragen mit IN/ (NOTIN) sollten auf der Grundlage der von der Datenbank maximal zulässigen Parameter aufgeteilt werden SQL
Backport	HIVE-20502: Fehler NPE beim Ausführen von skewjoin_mapjoin10.q behoben, wenn Spaltenstatistiken verwendet wurden.
Backport	HIVE-25765 : Die Eigenschaft skip.header.line.count überspringt Zeilen jedes Blocks, wenn die Datei größer ist FetchOperator
Fehler	Fehler NPE beim Einfügen in bestimmten Szenarien behoben, wenn beide aktiviert sind <code>hive.stats.column.autogather</code> <code>hive.groupby.skewindata</code>
Fehler	Problem beheben NPE, wenn kein <code>mapred.tasktracker.expiry.interval</code> Wert festgelegt ist

Amazon EMR 6.9.0 — Versionshinweise zu Hive

Amazon EMR 6.9.0 — Hive-Änderungen

Typ	Beschreibung
Upgrade	Aktualisieren von Jetty auf 9.4.48.v20220622
Upgrade	Unterstützung von Hadoop 3.3.3
Funktion	Amazon EMR Hive-Integration mit Lake Formation für interaktive Workloads mithilfe. GCSC API
Funktion	Amazon EMR Hive-Integration mit Iceberg.
Verbesserung	SSLIn HiveServer 2 aktivieren, wenn die Verschlüsselung während der Übertragung mithilfe der EMR Sicherheitskonfigurationen von Amazon aktiviert ist.
Verbesserung	Aktivieren Sie standardmäßig den optimierten Hive EMRFS Amazon S3 S3-Committer. Weitere Informationen finden Sie unter Aktivierung des optimierten Hive EMRFS S3-Committers .
Verbesserung	Fügen Sie hinzuHiveHBaseTableInputFormatV2 , dass nur die zugeordnete Version von InputFormat to fix -34210 erbt. SPARK Stellen Sie <code>hive.hbase.inputformat.v2</code> auf <code>true</code> ein, um es zu verwenden.
Verbesserung	Warten Sie, bis TezAM im Hintergrund mit hive.cli.tez.session.async gestartet wird, anstatt es zu beenden und sofort neu zu starten. Verwenden Sie diese <code>hive.emr.cli.tez.session.open.timeout</code> -Option, um dieses Timeout in Sekunden festzulegen.

Typ	Beschreibung
Verbesserung	Fügen Sie die Option hive.conf.restricted.list.append hinzu, um kommagetrennte Konfigurationen an die bestehende eingeschränkte Konfigurationsliste <code>hive.conf.restricted.list</code> anzuhängen.
Verbesserung	Klarere Fehlermeldung, wenn die Hive-Abfrage fehlschlägt, weil der Speicherort für die Datenbank nicht definiert ist.
Backport	HIVE-24484 : Aktualisieren Sie Hadoop auf 3.3.1 und Tez auf 0.10.2
Backport	HIVE-22398 : Entfernen Sie die Warteschlangenverwaltung über YARN ShimLoader
Backport	HIVE-23190 : LLAP: Ändern, um das Dateisystemobjekt IndexCache an zu übergeben. TezSpillRecord
Backport	HIVE-22185 : HADOOP -15832 wird Probleme bei Tests verursachen, die Cluster verwenden. MiniYarn
Backport	HIVE-21670 : Durch Abhängigkeit ersetzen. mockito-all mockito-core
Backport	HIVE-24542 : Guava für Upgrades vorbereiten.
Backport	HIVE-23751 QTest: Methode überschreiben <code>#mkdirs()</code> , um sie nach -16582 auszurichten. ProxyFileSystem HADOOP
Backport	HIVE-21603 : Java 11-Vorbereitung : Powermock-Version aktualisieren.

Typ	Beschreibung
Backport	HIVE-24083: hcatalog Fehler in Hadoop 3.3.0 : Authentifizierungstyp erforderlich.
Backport	HIVE-24282 : Spalten anzeigen sollte Ausgabespalten nicht sortieren, sofern nicht ausdrücklich anders angegeben.
Backport	HIVE-20656 : Sinnvolle Standardwerte: Speicherkonfigurationen für die Kartenaggregation sind zu aggressiv.
Backport	HIVE-25443 : Arrow SerDe kann komplexe Datentypen nicht serialisieren/deserialisieren, wenn es mehr als 1024 Werte gibt
Backport	HIVE-19792 : Führen Sie ein Upgrade von orc auf 1.5.2 durch und aktivieren Sie die Decimal_64-Schema-Evolutionstests.
Backport	HIVE-20437 : Behandelt die Schemaentwicklung von Float, Double und Decimal.
Backport	HIVE-21987 : Hive kann Parquet int32, das mit Dezimalzahlen annotiert ist, nicht lesen.
Backport	HIVE-20038 : Aktualisiere Abfragen für Tabellen ohne Buckets und partitionierte Threads. NPE

Amazon EMR 6.9.0 — Bekannte Probleme mit Hive

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um

das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR

Amazon EMR 6.8.0 — Versionshinweise zu Hive

Amazon EMR 6.8.0 — Hive-Änderungen

Typ	Beschreibung
Verbesserung	Reduzieren Sie die Dateisystemaufrufe im Befehl msck. Leistungsverbesserungen (~15-20x auf mehr als 10 000 Partitionen)
Backport	HIVE-20678 : HiveHBase TableOutputFormat sollte implementieren, um die Kompatibilität sicherzustellen HiveOutputFormat
Backport	HIVE-21040 : msck listet unnötige Dateien auf der letzten Ebene des Verzeichnisbaums auf
Backport	HIVE-21460 : Das Laden von Daten, gefolgt von einer Auswahl* -Abfrage, führt zu falschen Ergebnissen
Backport	HIVE-21660 : Falsches Ergebnis, wenn „Alle vereinigen“ und die spätere Ansicht mit Explode verwendet werden
Backport	HIVE-22505 : ClassCastException verursacht durch eine falsche Auswahl des vektorisierten Operators
Backport	HIVE-22513 : Die konstante Ausbreitung von gecasteten Spalten in Filteroperationen kann zu falschen Ergebnissen führen
Backport	HIVE-23435 : Beim vollständigen Outer-Join-Ergebnis fehlen Zeilen

Typ	Beschreibung
Backport	HIVE-24209 : Falsche Konvertierung der Suchargumente für den NOT BETWEEN Vorgang, wenn die Vektorisierung aktiviert ist
Backport	HIVE-24934 : In G Check ist keine Anmerkung erforderlich VectorizedExpressions enericUDF SQCount
Backport	HIVE-25278 : HiveProjectJoinTransposeRule Kann ungültige Transformationen mit Fensterausdrücken durchführen
Backport	HIVE-25505 : Falsche Ergebnisse mit Header. skip.header.line.count, wenn die erste Zeile leer ist
Backport	HIVE-26080 : Aktualisiere Accumulo-core auf 1.10.1
Backport	HIVE-26235 : Die OR-Bedingung für die Binärspalte gibt ein leeres Ergebnis zurück
Fehler	Korrigiert die Warnprotokolle mehrerer SLF4J Bindungen in stderr während des Starts
Fehler	Korrigiert die SHOW TABLE EXTENDED Abfrage, die mit dem Fehler Falscher FS fehlschlägt, wenn sich Partition und Tabelle auf unterschiedlichen Dateisystemen befinden.

Amazon EMR 6.8.0 — Bekannte Probleme mit Hive

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload

keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf `-1` zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR

Amazon EMR 6.7.0 — Versionshinweise zu Hive

Amazon EMR 6.7.0 — Hive-Änderungen

Typ	Beschreibung
Funktion	Amazon EMR Hive-Integration mit LakeFormation .
Funktion	Zusätzliche Auditprotokollierung für den optimierten Hive EMRFS Amazon S3 S3-Committer. Hive-Konfiguration: <code>hive.blobstore.output.committer.logging</code> , Standard: <code>false</code>
Funktion	Das Zielverzeichnis wurde beim Einfügen gelöscht und mit leerem Auswahlergebnis in eine unpartitionierte Tabelle/statische Partition überschrieben, sodass es sich ähnlich wie Hive 2.x verhält. Hive-Konfiguration: <code>hive.emr.iow.clean.target.dir</code> , Standard: <code>false</code>
Fehler	Es wurde ein zeitweiliger Abfragefehler behoben, der bei der Verwendung des optimierten Hive EMRFS Amazon S3 S3-Committers mit Partitions-Bucket-Sortierung auftrat.
Upgrade	Upgrade von Hive auf Version 3.1.3. Weitere Informationen finden Sie in den Versionshinweisen zu Apache Hive 3.1.3 .
Upgrade	Parquet wurde auf 1.12.2 aktualisiert .

Typ	Beschreibung
Backport	HIVE-20065 : Metastore sollte sich nicht auf Jackson 1.x verlassen
Backport	HIVE-20071 : Migrieren Sie zu Jackson 2.x und verhindern Sie die Nutzung
Backport	HIVE-20607 : TxnHandler sollte verwendet werden, um direkte Abfragen auszuführen PreparedStatement SQL
Backport	HIVE-20740 : Entfernen Sie den globalen Lock-In. ObjectStore setConf Methode
Backport	HIVE-20961: Implementierung einstellen NVL
Backport	HIVE-22059 : Hive-Exec-Jar enthält keine Jackson-Bibliothek (fasterxml)
Backport	HIVE-22351: Korrigiert die falsche Verwendung von Threads in ObjectStore TestObjectStore
Backport	HIVE-23534 : NPE in RetryingMetaStoreClient #invoke beim Fangen ohne Nachricht MetaException
Backport	HIVE-24048 : Harmonisieren Sie die Jackson-Komponenten auf Version 2.10.latest - Hive
Backport	HIVE-24768 : Verwende jackson-bom überall für den Versionsersatz
Backport	HIVE-24816 : Aktualisieren Sie Jackson aufgrund von -2020-25649 auf 2.10.5.1 oder 2.11.0+ CVE

Typ	Beschreibung
Backport	HIVE-25971: Das Herunterfahren der Tez-Task verzögert sich, weil der zwischengespeicherte Thread-Pool nicht geschlossen wurde
Backport	HIVE-26036 : verursacht durch () in NPE getMTable ObjectStore

Amazon EMR 6.7.0 — Bekannte Probleme mit Hive

- Abfragen mit Fensterfunktionen für dieselbe Spalte wie Join können zu ungültigen Transformationen führen, wie in [HIVE-25278](#) gemeldet, und zu falschen Ergebnissen oder Abfragefehlern führen. Eine Behelfslösung wäre, solche Abfragen auf CBO Abfrageebene zu deaktivieren. Das Update wird in einer EMR Amazon-Version nach 6.7.0 verfügbar sein. Für weitere Informationen wenden Sie sich an den AWS Support.
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR

Amazon EMR 6.6.0 — Versionshinweise zu Hive

Amazon EMR 6.6.0 — Hive-Änderungen

Typ	Beschreibung
Upgrade	Aktualisieren Sie Parquet auf 1.12.1
Upgrade	Aktualisieren der Jetty-Jar-Version auf 9.4.43.v20210629

Typ	Beschreibung
Fehler	Es wurde ein Problem behoben, das dazu führte, dass Hive auf allen Task-/Core-Knoten installiert wurde, wenn es auf einem Hive-Cluster aktiviert LLAP war.
Backport	HIVE-25942 : Aktualisieren Sie commons-io aufgrund von -2021-29425 auf 2.8.0 CVE
Backport	HIVE-25726: Upgrade-Geschwindigkeit auf 2.3 aufgrund von -2020-13936 CVE
Backport	HIVE-25680 : Autorisieren Sie #get_table_meta Server, um eines der Autorisierungsmodelle zu verwenden. HiveMetastore API HiveMetastore
Backport	HIVE-25554 : Aktualisieren Sie die Arrow-Version auf 0.15
Backport	HIVE-25242: Die Abfrage wird mit vectorize d.adaptor = chosen extrem langsam ausgeführt
Backport	HIVE-25085 : Clients werden nicht mehr sitzungsübergreifend gemeinsam genutzt . MetaStore
Backport	HIVE-24827 : Die Hive-Aggregationsabfrage gibt falsche Ergebnisse für Nicht-Textdateien zurück.
Backport	HIVE-24683 : Hadoop23Shims sind anfällig für nicht existierende Pfade getFileId NPE
Backport	HIVE-24656 : schlägt fehl bei Abfragen mit dem Wert Null bei Map- und Array-Typen CBO
Backport	HIVE-24556 : Optimiert für Fälle ohne Enkelkind DefaultGraphWalker

Typ	Beschreibung
Backport	HIVE-24408 : Parquet auf Version 1.11.1 aktualisieren
Backport	HIVE-24391 : Fehler in Branch-3.1 behoben FIX TestOrcFile
Backport	HIVE-24362 : Die AST Baumverarbeitung ist suboptimal für Bäume mit einer großen Anzahl von Knoten
Backport	HIVE-24316 : Upgrade ORC von 1.5.6 auf 1.5.8 in Branch-3.1
Backport	HIVE-24307 : Beeline mit Eigenschaftsdatei und -e-Parameter schlägt fehl
Backport	HIVE-24245 : Vektorisiert PTF mit „count“ und „distinct“ über die Partition, was zu falschen Ergebnissen führt.
Backport	HIVE-24224 : Das Überspringen von Kopf- und Fußzeilen für Hive on Tez in komprimierten Dateien wurde behoben
Backport	HIVE-24157 : Strikter Modus schlägt bei Timestamp fehl ↔ numerisch CAST
Backport	HIVE-24113 : in G NPE enericUDFTo UnixTimeStamp
Backport	HIVE-23987 : Aktualisieren Sie die Pfeilverion auf 0.11.0
Backport	HIVE-23972 : Fügt dem externen Client eine externe Client-ID hinzu LLAP

Typ	Beschreibung
Backport	HIVE-23806 : Vermeiden Sie es, den Status der Spaltenstatistiken in allen Partitionen zu löschen, falls das Schema erweitert wird. Dies verbessert die Laufzeit der Anweisung <code>alter table add columns</code> .
Backport	HIVE-23779 : BasicStatsTask Informationen werden in der Beeline-Konsole nicht gedruckt
Backport	HIVE-23306 : Der RESET Befehl funktioniert nicht, wenn vom System eine Konfiguration festgelegt wurde. <code>getProperty</code>
Backport	HIVE-23164 : Server wurde aufgrund von Threads, die keine Daemon sind, nicht richtig beendet
Backport	HIVE-22967 : Support <code>hive.reloadable.aux.jars.path</code> für Hive auf Tez
Backport	HIVE-22934 : Interaktive Protokollzähler des Hive-Servers zum Fehlerstream
Backport	HIVE-22901 : Das Ersetzen von Variablen kann dazu führen, dass keine Zirkelverweise entstehen OOM
Backport	HIVE-22769 : Falsche Abfrageergebnisse und Abfragefehler bei der Split-Generierung für komprimierte Textdateien
Backport	HIVE-22716 : Das Lesen in ist unterbrochen <code>ByteBuffer ParquetFooterInputFromCache</code>
Backport	HIVE-22648 : Parquet auf 1.11.0 aktualisieren

Typ	Beschreibung
Backport	HIVE-22640: Decimal64: wenn der Partitionsspaltenart Dezimal ist ColumnVector ClassCastException
Backport	HIVE-22621 : instabiler Testfall: TestLlapSignerImpl testSigning
Backport	HIVE-22533: Behebt mögliche LLAP Daemon-Web-UI-Schwachstellen
Backport	HIVE-22532 : PTFPPD Kann das Limit über die Funktion Rank/ falsch verschieben DenseRank
Backport	HIVE-22514 : HiveProtoLoggingHook könnte viel Speicher verbrauchen
Backport	HIVE-22476 : Die Hive-Datediff-Funktion lieferte inkonsistente Ergebnisse, wenn hive.fetch.task.conversion auf none gesetzt war
Backport	HIVE-22429 : Migrierte Clustertabellen mit bucketing_version 1 auf Hive 3 verwenden bucketing_version 2 für Einfügungen
Backport	HIVE StatsUtils NPE-22412 : Wird bei der Erklärung ausgelöst
Backport	HIVE-22360 : MultiDelimitSerDe gibt falsche Ergebnisse in der letzten Spalte zurück, wenn die geladene Datei mehr Spalten hat als die im Tabellenschema
Backport	HIVE-22332 : Hive sollte sicherstellen, dass die Einstellungen für die Schemaentwicklung seit -540 gültig sind ORC

Typ	Beschreibung
Backport	HIVE-22331 : unix_timestamp ohne Argument gibt den Zeitstempel in Millisekunden statt in Sekunden zurück
Backport	HIVE OperationManager-22275 : queryIdOperation bereinigt mehrere nicht richtig queryIds
Backport	HIVE-22273 : Die Zugriffsprüfung schlägt fehl, wenn ein temporäres Verzeichnis entfernt wird
Backport	HIVE-22270 : Aktualisieren Sie commons-io auf 2.6
Backport	HIVE-22241 : Implementierung zur Interpretation von Datum/Zeitstempel unter Verwendung seiner UDF internen Darstellung und des gregorianisch-julianischen Hybridkalenders
Backport	HIVE-22241 : Implementierung zur Interpretation von Datum/Zeitstempel unter Verwendung seiner internen Darstellung und des UDF gregorianisch-julianischen Hybrids
Backport	HIVE-22232 : wenn hive.order.columnalignment auf false gesetzt ist NPE
Backport	HIVE-22231 : Hive-Abfrage mit großer Größe über Knox schlägt fehl, Broken Pipe Write ist fehlgeschlagen
Backport	HIVE-22221 : Externer Llap-Client — # muss reduziert werden LlapBaseInputFormat getSplits

Typ	Beschreibung
Backport	HIVE-22208 : Der Spaltenname mit reserviertem Schlüsselwort wird nicht maskiert, wenn die Abfrage, die eine Verknüpfung für eine Tabelle mit Maskenspalte enthält, neu geschrieben wird
Backport	HIVE-22197 : Ein häufiger Merge-Join, der eine Klassenumwandlungs-Ausnahme auslöst.
Backport	HIVE-22170 : from_unixtime und unix_time stamp sollten die Zeitzone der Benutzersitzung verwenden
Backport	HIVE-22169 : Tez: versucht nach Plandateien zu suchen, die für Tez nicht existieren SplitGenerator
Backport	HIVE-22168 : Entferne sehr teures Logging aus dem Llap-Cache-Hotpath
Backport	HIVE-22161 : Synchronisiert auf org.apache.hadoop.hive.ql.udf. UDF FunctionRegistry UDFTypeKlasse
Backport	HIVE-22120 : Korrigiert falsche ArrayOutOfBounds Ergebnisse/Ausnahmen bei Verbindungen auf der linken äußeren Karte unter bestimmten Randbedingungen
Backport	HIVE-22115 : Verhindert die Erstellung eines Query Routing Appenders, wenn die Eigenschaft auf „Falsch“ gesetzt ist
Backport	HIVE-22113 : Verhindert das Herunterfahren bei verwandten LLAP AMReporter RuntimeException

Typ	Beschreibung
Backport	HIVE-22106 : Abfrageübergreifende Synchronisation für die Partition-Eval entfernen
Backport	HIVE-22099 : Seit -2007 UDFs können mehrere datumsbezogene Daten nicht richtig mit julianischen Daten umgehen HIVE
Backport	HIVE-22037 : HS2 sollte sich protokollieren, wenn das System heruntergefahren wird aufgrund von OOM
Backport	HIVE-21976 : Der Offset sollte in Calcit Null statt Null sein HiveSortLimit
Backport	HIVE-21924 : Textdateien teilen, auch wenn Kopf-/Fußzeile vorhanden ist
Backport	HIVE-21913 : GenericUDTFGet Splits sollte Benutzernamen genauso behandeln wie LLAP
Backport	HIVE-21905 : Verbesserung der Generika in der gesamten Klasse FetchOperator
Backport	HIVE-21902 : HiveServer 2 UI: Der Jetty-Antwort-Header benötigt X-Frame-Options
Backport	HIVE-21888: Setze hive.parquet.timesamp.skip.conversion standardmäßig auf true
Backport	HIVE-21868 CAST: Vektorisieren... FORMAT
Backport	HIVE-21864 : # LlapBaseInputFormat closeAll
Backport	HIVE-21863 : Verbessert die Umwandlung des Vectorizer-Typs für Expression WHEN
Backport	HIVE-21862 : ORC ppd erzeugt ein falsches Ergebnis mit Zeitstempel

Typ	Beschreibung
Backport	HIVE-21846 : Erstelle einen Thread in TeamAm, der regelmäßig Metriken abrufft LlapDaemon
Backport	HIVE-21837 : löst eine Ausnahme aus, wenn MapJoin die ausgewählte Spalte komplett Nullwerte hat
Backport	HIVE-21834 : Vermeiden Sie unnötige Aufrufe, um die Filterbedingungen zu vereinfachen
Backport	HIVE-21832 : Neue Metriken zur Ermittlung der durchschnittlichen Warteschlangen-/Servier-/Antwortzeit
Backport	HIVE-21827 : Mehrere eingehende Aufrufe durchlaufen die Methode nicht SemanticAnalyzer.getTableObject ByName
Backport	HIVE-21822 : Machen Sie LlapDaemon Metriken mit einer neuen Methode verfügbar API
Backport	HIVE-21818CBO : Beim Kopieren kommt es zu Metastore-Verkehr TableRelOptHiveTable
Backport	HIVE-21815 : Statistiken in der Datei werden zweimal analysiert ORC
Backport	HIVE-21805:2: HiveServer Benutze das Schnelle ShutdownHookManager APIs
Backport	HIVE-21799 : NullPointerException in DynamicPartitionPruningOptimization, wenn sich der Join-Schlüssel in der Aggregationspalte befindet

Typ	Beschreibung
Backport	HIVE-21794 : Fügt Parameter für materialisierte Ansichten hinzu zu sqlStdAuth SafeVarNameRegexes
Backport	HIVE-21768:JDBC: Entfernt das Standard-Union-Präfix für nicht eingeschlossene Abfragen UNION
Backport	HIVE-21746 : ArrayIndexOutOfBoundsException während eines dynamisch partitionierten Hash-Joins, bei deaktivierter Option CBO
Backport	HIVE-21717 : Das Umbenennen des Verzeichnisses in der Verschiebungsaufgabe schlägt fehl.
Backport	HIVE-21685 : Falsche Vereinfachung in der Abfrage mit mehreren IN-Klauseln
Backport	HIVE-21681 : Im Format Describe werden falsche Informationen für mehrere Primärschlüssel angezeigt
Backport	HIVE-21651 : Verschiebe Protobug serde nach Hive-Exec.
Backport	HIVE-21619 : Gibt den Zeitstempeltyp ohne Genauigkeit in Explain Extended aus SQL
Backport	HIVE-21592 : OptimizedSql wird nicht angezeigt , wenn der Ausdruck Folgendes enthält CONCAT
Backport	HIVE-21576 : Stellt vor... CAST FORMATund begrenzte Liste SQL von:Datums-/Uhrzeitformaten für 2016

Typ	Beschreibung
Backport	HIVE-21573 : Der binäre Transport ignoriert den Principal, wenn auth auf gesetzt ist delegationToken
Backport	HIVE-21550 : TestObjectStore Tests sind fehlerhaft — Eine Sperre konnte nicht innerhalb der angeforderten Zeit abgerufen werden
Backport	HIVE-21544 : Konstante Ausbreitung verfälscht Koalesce/Case/When Ausdrücke beim Falten
Backport	HIVE-21539: + WHERE-Klausel in derselben Spalte führt zu einem falschen Umschreiben der Abfrage GroupBy
Backport	HIVE-21538 : Beeline: Passwortquelle, obwohl der Konsolenleser den Verbindungsparameter nicht übergeben hat
Backport	HIVE-21509 : LLAP Kann beschädigte Spaltenvektoren zwischenspeichern und ein falsches Abfrageergebnis zurückgeben
Backport	HIVE-21499 : sollte die Funktion nicht aus der Registrierung entfernen, wenn der Befehl create fehlschlug mit AlreadyExistsException
Backport	HIVE-21496 : Bei der automatischen Dimensionierung ungeordneter Puffer kann es zu einem Überlauf kommen
Backport	HIVE-21468 : Berücksichtigung von Groß- und Kleinschreibung bei Bezeichnernamen für Speicherhandler JDBC
Backport	HIVE-21467 : Entfernen Sie veraltete JUnit.Framework.Assert-Importe

Typ	Beschreibung
Backport	HIVE-21435 : LlapBaselInputFormat sollte beim Erstellen die Aufgabennummer aus __ID conf abrufen, falls vorhanden TASK ATTEMPT SubmitWorkRequestProto
Backport	HIVE-21389 : Bei der Hive-Distribution fehlt javax.ws.rs-api.jar nach -21247 HIVE
Backport	HIVE-21385 : Erlaubt die Deaktivierung der Übertragung von Berechnungen, die nicht splittbar sind, an Quellen JDBC
Backport	HIVE-21383: JDBC Speicherhandler: Benutzt Katalog und Schema, um Tabellen abzurufen, falls angegeben
Backport	HIVE-21382 : Optimierung der Reduzierung der Gruppierung nach Schlüsseln — Schlüssel werden in query23 nicht reduziert
Backport	HIVE-21362 : Fügt ein Eingabeformat und serde hinzu, um aus Protobuf-Dateien zu lesen.
Backport	HIVE-21340 :: CBO Schneidet Spalten, die keine Schlüssel sind, aus SemiJoin
Backport	HIVE-21332 : Löscht die nicht gesperrten Puffer statt der gesperrten
Backport	HIVE-21329 : Ungeordnete Größe des Ausgabepuffers zur benutzerdefinierten Tez-Laufzeit, abhängig von der Operator-Pipeline
Backport	HIVE-21295 : StorageHandler soll das Datum unter Verwendung der Hive-Konvention in eine Zeichenfolge konvertieren

Typ	Beschreibung
Backport	HIVE-21294 : Vektorisierung: 1-Reducer Shuffle kann die Objekt-Hash-Funktionen überspringen
Backport	HIVE QueryConditionBuilder -21255 : In entfernen JdbcStorageHandler
Backport	HIVE-21253: Support in DB2 JDBC StorageHandler
Backport	HIVE-21232LLAP : Fügt einen Cache-Miss-freundlichen Split-Affinitätsanbieter hinzu
Backport	HIVE-21214 : MoveTask : Wird attemptId anstelle der Dateigröße für die Deduplizierung von Dateien verwendet compareTempOr DuplicateFiles
Backport	HIVE-21184 : Fügt einen Plan im Format „Erläutern und Erläutern“ mit Kosteninformationen hinzu CBO
Backport	HIVE-21182 : Überspringe die Einrichtung von Hive Scratch Dir während der Planung
Backport	HIVE-21171 : Überspringe das Erstellen von Scratch-Verzeichnissen für Tez, wenn es aktiviert ist RPC
Backport	HIVE-21126 : Abfragen auf Sitzungsebene in # zulassen LlapBaseInputFormat getSplit
Backport	HIVE-21107 : Fehler „Feld kann nicht gefunden werden“ bei dynamisch partitioniertem Hash-Join
Backport	HIVE-21061 : Die CTAS Abfrage schlägt fehl, wenn die Quelle leer ist IllegalStateException

Typ	Beschreibung
Backport	HIVE-21041 : NPE, beim Abrufen des Schemas aus ParseException dem logischen Plan
Backport	HIVE-21013 : Die Partitionsspalte JdbcStorageHandler konnte in Oracle nicht gefunden werden
Backport	HIVE-21006 : Erweitern, um Semijoin SharedWorkOptimizer zu entfernen, wenn sich eine Möglichkeit zur Wiederverwendung ergibt
Backport	HIVE-20992 : Teile die Konfiguration hive.metastore.dbaccess.ssl.properties in aussagekräftigere Konfigurationen auf
Backport	HIVE-20989 JDBC: - GetOperationStatus Das +-Protokoll kann den Abfragefortschritt im Ruhezustand blockieren
Backport	HIVE-20988 : Falsche Ergebnisse für Gruppierungsabfragen mit Primärschlüsseln für mehrere Spalten
Backport	HIVE-20985 : Wenn es sich bei den Eingaben des Select-Operators um temporäre Spalten handelt, kann die Vektorisierung einige davon als Ausgabe wiederverwenden
Backport	HIVE-20978 : „hive.jdbc.*“ sollte Folgendes hinzufügen sqlStdAuth SafeVarNameRegexes
Backport	HIVE-20953 : Löscht eine Funktion aus der Funktionsregistrierung, wenn sie beim Erstellen nicht zum Metastore hinzugefügt werden kann.
Backport	HIVE-20952 : .java reinigen VectorizationContext

Typ	Beschreibung
Backport	HIVE-20951 :: Setze Xms immer auf 50% LLAP
Backport	HIVE-20949 : Verbessert die Schätzung der Kardinalität bei der Raumplanung PKFK
Backport	HIVE-20944 : Statistiken werden während der Abfragekompilierung nicht validiert
Backport	HIVE-20940 : Überbrückt Fälle, in denen die Typauflösung von Calcite strenger ist als bei Hive.
Backport	HIVE-20937 : Die Postgres-JDBC-Abfrage schlägt mit „darf nicht negativ sein“ fehl LIMIT
Backport	HIVE-20926 : Der Hinweis zur Reduzierung von Semi-Joins schlägt fehl, wenn Bloom-Filter-Einträge hoch sind oder wenn es keine Statistiken gibt
Backport	HIVE-20920 : Verwenden Sie SQL Einschränkungen, um den Algorithmus zur Neuordnung von Verknüpfungen zu verbessern
Backport	HIVE-20918 : Markierung zum Aktivieren/Deaktivieren der Übertragung von Berechnungen von Calcite in die Verbindung JDBC
Backport	HIVE-20915 : Stellt die dynamische Optimierung der Sortierpartition für HoS und MR zur Verfügung
Backport	HIVE-20910 : Das Einfügen in eine Bucket-Tabelle schlägt aufgrund der Optimierung der dynamischen Partitionssortierung fehl

Typ	Beschreibung
Backport	HIVE-20899 : Keytab URI for LLAP YARN Service ist nur auf Support beschränkt HDFS
Backport	HIVE-20898 : Für zeitbezogene Funktionen dürfen Argumente nicht in einen Typ umgewandelt werden, der keine NULL-Werte zulässt
Backport	HIVE-20881 : Konstante Ausbreitung vereinfacht Prognosen zu stark
Backport	HIVE-20880 : Den Standardwert für hive.stats.filter.in.min.ratio aktualisieren
Backport	HIVE-20873 : Verwenden Sie VectorHashKeyWrapperTwoLong Murmur-Hash, um die Hash-Kollision zu reduzieren
Backport	HIVE-20868 : SMB Join schlägt zeitweise fehl, wenn ein untergeordnetes Op-In ist TezDummyOperator getFinalOp MapRecord Processor
Backport	HIVE-20853 : Enthüllen . ShuffleHandler registerDagim Schoßdämon API
Backport	HIVE-20850 : Wenn möglich, wird die Groß- und Kleinschreibung von Projektionen in Dimensionstabellen übertragen
Backport	HIVE-20842 : Die in HIVE -20660 eingeführte Logik zur Schätzung von Statistiken für Gruppen nach wurde korrigiert
Backport	HIVE-20839 : Fehler „Feld kann nicht gefunden werden“ beim dynamisch partitionierten Hash-Join

Typ	Beschreibung
Backport	HIVE-20835 : Die Interaktion zwischen Einschränkungen und MV-Umschreiben kann zu einer Schleife im Calcite-Planer führen
Backport	HIVE-20834 : QueryResultCache Hive-Einträge, die auf die zwischengespeicherte Abfrage verweisen SemanticAnalyzer
Backport	HIVE-20830 : In einigen Fällen schlägt die Bestätigung der JdbcStorageHandler Bereichsabfrage fehl
Backport	HIVE-20829 : Würfe mit Bereichsaufteilung JdbcStorageHandler NPE
Backport	HIVE-20827 : Inkonsistente Ergebnisse für leere Arrays
Backport	HIVE-20826 : Die HiveSemiJoin Regel wurde erweitert, sodass Join + Group by auf der linken Seite in Left Semi Join umgewandelt wird
Backport	HIVE-20821 : Schreiben Sie 0 in eine +- Kombination SUM um SUM COALESCE
Backport	HIVE-20815 : JdbcRecordReader .next darf keine Ausnahme essen
Backport	HIVE-20813 : udf to_epoch_milli muss auch Timestamp ohne Zeitzone unterstützen.
Backport	HIVE-20804: Weitere Verbesserungen bei der Gruppierung durch Optimierung mit Einschränkungen
Backport	HIVE-20792 : Beim Einfügen eines Zeitstempels mit Zonen werden die Daten gekürzt

Typ	Beschreibung
Backport	HIVE-20788 : Durch die erweiterte SJ-Reduzierung werden Spalten beim Erstellen von Filtern möglicherweise falsch zurückverfolgt
Backport	HIVE-20778 : Die Neuordnung von Verknüpfungen wird möglicherweise nicht ausgelöst, wenn alle Joins im Plan durch die Dekorrelationslogik erstellt wurden
Backport	HIVE-20772 : Zähler pro Aufgabe aufzeichnen in CPU LLAP
Backport	HIVE-20768: Tumbling Window wird hinzugefügt UDF
Backport	HIVE-20767 : Mehrere Projekte zwischen Join-Operatoren können sich auf die Neuordnung von Verknüpfungen mithilfe von Einschränkungen auswirken
Backport	HIVE-20762 : Das LOG Bereinigungsintervall ist mit 60 NOTIFICATION Sekunden fest codiert und ist zu klein
Backport	HIVE-20761 : In der notification_sequence-Tabelle zur Aktualisierung ausgewählt, das Wiederholungsintervall und die Anzahl der Wiederholungen ist zu klein
Backport	HIVE-20751 : Aktualisieren Sie die Pfeilverision auf 0.10.0
Backport	HIVE-20746 : Schließt die Datei am Ende des HiveProtoHookLogger Tages nicht.

Typ	Beschreibung
Backport	HIVE-20744 : Verwenden SQL Sie Einschränkungen, um den Algorithmus zur Neuordnung von Verknüpfungen zu verbessern
Backport	HIVE-20740 : Entfernen Sie die globale Sperre. ObjectStore setConf Methode. Dieser CherryPick portiert HIVE -20740, der für Hive 3.2 und 4.x vorgesehen war, auf 3.1.x zurück
Backport	HIVE-20734 : Beeline: Wenn beeline-site.xml ist und Hive zu Beeline CLI weiterleitet, sollte es den Systembenutzernamen/das Dummy-Passwort verwenden, anstatt nach einem zu fragen
Backport	HIVE-20731 : Die Eingabe der Keystore-Datei sollte autorisiert sein JdbcStorageHandler
Backport	HIVE-20720 : Fügt dem Handler eine Partitionsspaltenoption hinzu JDBC
Backport	HIVE-20719 : Die SELECT Anweisung schlägt fehl, wenn die Optimierung und Vektorisierung von hive.optimize.sort.dynamic.partition UPDATE aktiviert ist
Backport	HIVE-20718 : Perf-CLI-Treiber mit Einschränkungen hinzugefügt
Backport	HIVE-20716 : Setze den Standardwert für hive.cbo.stats.correlated.multi.key.join auf true
Backport	HIVE HivePointLookupOptimizer -20712 : sollte tiefe Fälle extrahieren
Backport	HIVE-20710 : Konstantes Folding erzeugt möglicherweise keine Nullkonstanten ohne Typen

Typ	Beschreibung
Backport	HIVE-20706 : external_jdbc_table2.q schlägt zeitweise fehl
Backport	HIVE-20704 : Erweitern, um andere Funktionen zu unterstützen HivePreFilteringRule
Backport	HIVE-20703 : Unterstellen Sie die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung
Backport	HIVE-20702 : Berücksichtigen Sie den Mehraufwand, der durch datenstrukturabhängige Schätzungen bei der Mapjoin-Auswahl entsteht
Backport	HIVE-20692 : Aktiviert die Faltung von x IS () [[]]-Ausdrücken NOT NOT TRUE FALSE
Backport	HIVE-20691 : Korrigieren Sie org.apache.hadoop.hive.cli. TestMiniLlapCliDriver. testCliDriver[cttl]
Backport	HIVE-20682 : Die asynchrone Abfrageausführung kann möglicherweise fehlschlagen, wenn Shared vom Master-Thread geschlossen wird sessionHive
Backport	HIVE-20676 : HiveServer 2: ist nicht auf den Daemon-Status gesetzt PrivilegeSynchronizer
Backport	HIVE-20660 : Die Schätzung der Gruppierung nach Statistiken könnte verbessert werden, indem die Gesamtzahl der Zeilen auf die Quelltable begrenzt wird

Typ	Beschreibung
Backport	HIVE-20652 : JdbcStorageHandler Push-Join von zwei verschiedenen Datenquellen zum JDBC-Treiber
Backport	HIVE-20651 : JdbcStorageHandler Das Passwort sollte verschlüsselt sein
Backport	HIVE-20649 : LLAP bewusster Speichermanager für Orc-Autoren
Backport	HIVE-20648 LLAP: Vektorgruppe nach Operator sollte Speicher pro Executor verwenden
Backport	HIVE-20646 : Die Partitionsfilter-Bedingung wird nicht auf die Metastore-Abfrage heruntergeschoben, wenn sie IS hat NOT NULL
Backport	HIVE-20644 : Vermeiden Sie es, vertrauliche Informationen durch eine Hive-Runtime-Ausnahme preiszugeben
Backport	HIVE-20636 : Verbessert die Schätzung der Anzahl der Nullwerte nach dem Outer-Join
Backport	HIVE-20632 : Die Abfrage mit get_splits UDF schlägt fehl, wenn eine materialisierte Ansicht für die abgefragte Tabelle erstellt wird
Backport	HIVE-20627 : Gleichzeitige asynchrone Abfragen schlagen zeitweise fehl und verursachen Speicherleck LockException
Backport	HIVE-20623 : Gemeinsames Arbeiten: Erweitert die gemeinsame Nutzung von Map-Join-Cache-Einträgen in LLAP

Typ	Beschreibung
Backport	HIVE-20619 : Standardmäßig in 2 einschließen MultiDelimitSerDe HiveServer
Backport	HIVE-20618 : Beim Join BucketMapJoin kann die Auswahl für Tabellen ohne Buckets ausgewählt werden
Backport	HIVE-20617 : Korrigiert den Typ der Konstanten in IN-Ausdrücken, sodass sie den richtigen Typ haben
Backport	HIVE-20612 : Erstellt ein neues Join-Korrelationsflag mit mehreren Schlüsseln für CBO
Backport	HIVE-20603 : Fehler „Falsches FS“ beim Einfügen in eine Partition nach dem Ändern des Dateisystems mit dem Speicherort der Tabelle
Backport	HIVE-20601 : Null in _ Ereignis in EnvironmentContext ALTER PARTITION DbNotificationListener
Backport	HIVE-20583 : Verwenden Sie den kanonischen Hostnamen nur für die Kerberos-Authentifizierung HiveConnection
Backport	HIVE-20582 : Hflush im Hive-Proto-Logging konfigurierbar machen
Backport	HIVE-20563 : Vektorisierung: Der Ausdruck schlägt fehl, wenn der Typ/und der Ergebnistyp unterschiedlich sind CASE WHEN THEN ELSE
Backport	HIVE-20558 : Ändern Sie den Standardwert von hive.hashtable.key.count.adjustment auf 0,99

Typ	Beschreibung
Backport	HIVE-20552 LogicalPlan: Schema schneller abrufen
Backport	HIVE-20550 : Wechsle W, um Beeline zum Senden von ebHCat Hive-Abfragen zu verwenden
Backport	HIVE-20537 : In mehreren Spalten werden Schätzungen mit unkorrelierten Spalten verknüpft, die sich in und Hive unterscheiden CBO
Backport	HIVE-20524 : Die Überprüfung der Schemaentwicklung ist beim Übergang von Hive-Version 2 zu Version 3 für bis unterbrochen ALTER TABLE VARCHAR DECIMAL
Backport	HIVE-20522 : HiveFilterSetOpTransposeRule Kann aufgrund der NULL-Zulässigkeit von Feldern einen Assertionsfehler auslösen
Backport	HIVE-20521 : HS2 doAs =true hat ein Berechtigungsproblem mit hadoop.tmp.dir, mit MR- und S3A-Dateisystemen
Backport	HIVE-20515 : Leere Abfrageergebnisse, wenn der Ergebnis-Cache und das temporäre Abfrageverzeichnis verwendet werden, das Ergebnis-Cache-Verzeichnis in verschiedenen Dateisystemen
Backport	HIVE-20508 : Hive unterstützt keine Benutzernamen vom Typ „user @realm“
Backport	HIVE-20507 : Beeline : Fügt einen Utility-Befehl hinzu, um alle Uris aus beeline-site.xml abzurufen

Typ	Beschreibung
Backport	HIVE-20505: aktualisiere org.openjdk.jmh:jmh-core auf Version 1.21
Backport	HIVE-20503 : Bei der Mapjoin-Auswahl werden Schätzungen verwendet, die die Datenstruktur berücksichtigen
Backport	HIVE-20498 : Support den Datumstyp für die automatische Erfassung von Spaltenstatistiken
Backport	HIVE-20496 : Vektorisierung: Vektorisiert PTF IllegalStateException
Backport	HIVE-20494 : G ist nach -19440 kaputt enericUDFRestrict InformationSchema HIVE
Backport	HIVE-20477 : wird nicht angezeigt, wenn der Ausdruck Folgendes enthält OptimizedSql INs
Backport	HIVE-20467 : Erlaube IF NOT EXISTS /IF bei der Erstellung/Löschung eines Resource nplans EXISTS
Backport	HIVE-20462 : "IF" CREATE VIEW schlägt fehl, wenn die Ansicht bereits existiert NOT EXISTS
Backport	HIVE-20455 : Log-Ausgabe von security.authorization.PrivilegeSynchronizer.ausführen
Backport	HIVE-20439 : Benutzt das überhöhte Speicherlimit bei der Join-Auswahl für Llap
Backport	HIVE-20433 : Die implizite Konvertierung von Zeichenkette in Zeitstempel ist langsam

Typ	Beschreibung
Backport	HIVE-20432 : Für Integer-Typen zur Schätzung von Statistiken BETWEEN nach IN umschreiben
Backport	HIVE-20423 : Als Standard-Nullreihenfolge festlegen NULLS LAST
Backport	HIVE-20418 : LLAP IO verarbeitet ORC Dateien, bei denen der Zeilenindex deaktiviert ist, bei Abfragen, bei denen keine Spalten ausgewählt sind, möglicherweise nicht korrekt
Backport	HIVE-20412 : ein NPE HiveMetaHook
Backport	HIVE-20406 : Nested Coalesce liefert falsche Ergebnisse
Backport	HIVE-20399 : Wenn CTAS ein benutzerdefinierter Tabellenspeicherort nicht vollständig qualifiziert ist, schlägt er bei MM-Tabellen fehl
Backport	HIVE-20393 : Semijoin Reduction : For verhält sich inkonsistent markSemiJoin DPP
Backport	HIVE-20391 : HiveAggregateReduceFunction sRule kann bei der Zerlegung der Aggregatfunktion auf einen falschen Rückgabotyp schließen
Backport	HIVE-20383 : Ungültiger Warteschlangenname und Synchronisationsprobleme im Hive Proto Events Hook.
Backport	HIVE-20367 : Vektorisierung: Support Streaming für,,, PTF AVG MAX MIN SUM

Typ	Beschreibung
Backport	HIVE-20366 : Die Schätzungen der Statistiken von TPC -DS query78 sind für seinen Nullfilter deaktiviert
Backport	HIVE-20364 : Aktualisiere die Standard-einstellung für <code>hive.map.aggr.hash.min.reduction</code>
Backport	HIVE-20352 : Vektorisierung: Gruppierungsfunktion Support
Backport	HIVE-20347 : <code>hive.optimize.sort.dynamic.partition</code> sollte mit <code>partitioned</code> und <code>MV</code> funktionieren CTAS
Backport	HIVE-20345 : Die Drop-Datenbank kann hängen bleiben, wenn die Tabellen bei einem anderen Aufruf gelöscht werden
Backport	HIVE-20343 : Hive 3: respektiert <code>transactional_properties</code> nicht CTAS
Backport	HIVE-20340 : Druid Needs Explicit CASTs von Timestamp bis zu dem Zeitpunkt, zu dem die Ausgabe der Timestamp-Funktion als Zeichenfolge verwendet wird STRING
Backport	HIVE-20339 : Vektorisierung: Hebt unnötige Einschränkungen auf, sodass manche Einschränkungen nicht vektorisiert werden PTF RANK
Backport	HIVE-20337 :: <code>CachedStore Expr</code> füllt die Partitionsliste nicht korrekt <code>getPartitionsBy</code>
Backport	HIVE-20336 : Maskierungs- und Filterrichtlinien für materialisierte Ansichten

Typ	Beschreibung
Backport	HIVE-20326 : Erstellen Sie Einschränkungen mit der Standardeinstellung anstelle von NEIN RELY RELY
Backport	HIVE-20321 : Vektorisierung: Reduziert die Speichergröße von 1 Spalte auf <1 VectorHas hKeyWrapper CacheLine
Backport	HIVE-20320 : Aktiviere das Flag hive.optimize.remove.sq_count_check
Backport	HIVE-20315 : Vektorisierung: Behebt weitere Probleme mit falschen Ergebnissen und vermeidet unnötige Umwandlungen/Konvertierungen NULL
Backport	HIVE-20314 : Schließt das Bereinigen von Partitionen in das Umschreiben materialisierter Ansichten ein
Backport	HIVE-20312 : Erlaubt es Arrow-Clients, ihre eigenen zu verwenden mit BufferAllocator LlapOutputFormatService
Backport	HIVE-20302 :LLAP: Die nicht vektorisierte Ausführung in IO ignoriert virtuelle Spalten, einschließlich __ID ROW
Backport	HIVE-20300 : VectorFileSinkArrowOperator
Backport	HIVE-20299 : potenzielles Rennen im Signer-Unit-Test LLAP
Backport	HIVE-20296 : Verbessern, HivePointLookupOptimizerRule um Extrahieren aus komplexeren Kontexten zu ermöglichen

Typ	Beschreibung
Backport	HIVE-20294 : Vektorisierung: Probleme mit / Falschen Ergebnissen in/behoben NULL COALESCE ELT
Backport	HIVE-20292 : Schlechte Join-Reihenfolge in tpchds query93 mit definierter primärer Einschränkung
Backport	HIVE-20290 : Lazy initialisiert, sodass währenddessen keine Puffer zugewiesen werden ArrowColumnarBatchSerDe GetSplits
Backport	HIVE-20281 : SharedWorkOptimizer schlägt fehl, wenn „der Inhalt des Operator-Cache und der tatsächliche Plan unterschiedlich sind“
Backport	HIVE-20277 : Vektorisierung: Groß - und Kleinschreibung, die zurückgegeben werden, werden nicht unterstützt für BOOLEAN FILTER
Backport	HIVE-20267 : Erweiterung der WebUI um ein Formular zur dynamischen Konfiguration von Log-Levels
Backport	HIVE-20263 : Tippfehler in der Variablen HiveReduceExpressionsWithStatsRule
Backport	HIVE-20260 : NDV einer Spalte sollte nicht skaliert werden, wenn die Zeilenanzahl durch einen Filter in einer anderen Spalte geändert wird
Backport	HIVE-20252 : Reduzierung von Semijoins: Zyklen aufgrund von Semi-Join-Verzweigungen können unentdeckt bleiben, wenn der kleinen Tabellenseite ein Map-Join vorgeschaltet ist.

Typ	Beschreibung
Backport	HIVE-20245 : Vektorisierung: Probleme mit// Falschen Ergebnissen in /IN behoben NULL BETWEEN
Backport	HIVE-20241 : Support Partitionierungsspezifikation in Anweisungen CTAS
Backport	HIVE-20240 : Semijoin Reduction: Verwenden Sie die lokale Variable, um den Zustand der externen Tabelle zu überprüfen
Backport	HIVE-20226 : löst eine Ausnahme aus, wenn HMS getNextNotification die Anfrage die max_rows der Tabelle überschreitet maxEvents
Backport	HIVE-20225 : zur Unterstützung des Teradata-Binärformats SerDe
Backport	HIVE-20213 : Aktualisieren Sie Calcite auf 1.17.0
Backport	HIVE-20212 : Hiveserver2 gibt im HTTP-Modus die Metrik Default.General.OPEN_CONNECTIONS falsch aus
Backport	HIVE-20210 : Der Simple Fetch-Optimizer sollte dazu führen, dass nach Spalten gefiltert wird, die nicht partitioniert sind und die Konvertierung minimal ist MapReduce
Backport	HIVE-20209 : Die Metastore-Verbindung schlägt beim ersten Versuch im Repl-Dump fehl
Backport	HIVE-20207 : Vektorisierung: Behebt// Probleme mit falschen Ergebnissen beim Filtern// Vergleichen NULL

Typ	Beschreibung
Backport	HIVE-20204 : Typkonvertierung bei IN
Backport	HIVE-20203 : Bei Arrow gibt es ein unbekanntes SerDe DirectByteBuffer
Backport	HIVE-20197: Vektorisierung: Fügt DECIMAL _64-Tests hinzu, fügt Datums-/Interval-/Zeitstempel-Arithmetik hinzu und fügt weitere BY-Aggregation hinzu GROUP
Backport	HIVE cboInfo -20193 : ist im Explain-Plan-Json nicht vorhanden
Backport	HIVE-20192 : HS2 Bei eingebettetem Metastore sind Objekte undicht JDOPersistenceManager
Backport	HIVE-20183 : Das Einfügen aus einer Bucket-Tabelle kann zu Datenverlust führen, wenn die Quelltable einen leeren Bucket enthält
Backport	HIVE-20177 : Vektorisierung: Reduziert die Zuweisung im Streaming-Modus KeyWrapper GroupBy
Backport	HIVE-20174 : Vektorisierung: Probleme mit / falschen Ergebnissen in BY-Aggregationsfunktionen behoben NULL GROUP
Backport	HIVE-20172 : Beim Versuch, eine Verbindung zum StatsUpdater Remote-Metastore herzustellen, ist GSS eine Ausnahme aufgetreten
Backport	HIVE-20153 : Count und Sum UDF verbrauchen mehr Speicher in Hive 2+

Typ	Beschreibung
Backport	HIVE-20152 : setzt den Datenbankstatus zurück, wenn der Repl-Dump fehlschlägt, sodass die Tabelle umbenannt werden kann
Backport	HIVE-20149: Fehlgeschlagen/Timeout TestHiveCli
Backport	HIVE-20130 : Bessere Protokollierung für den Informationsschema-Synchronizer
Backport	HIVE-20129 : Kehren Sie zur positionsbasierten Schemaentwicklung für Orc-Tabellen zurück
Backport	HIVE-20118 : SessionStateUserAuthenticator. getGroupNames
Backport	HIVE-20116 : verwendet den TezTask übergeordneten Logger
Backport	HIVE-20115 : Bei Säuretabellen sollte kein Fußzeilen-Scan zur Analyse verwendet werden
Backport	HIVE-20103 : WM: DAG Zähler nur aggregieren, wenn mindestens einer verwendet wird
Backport	HIVE-20101 : BloomKFilter: Vermeide es, die lokalen Byte [] -Arrays vollständig zu verwenden
Backport	HIVE-20100 : OpTraits : Select Optraits sollte aufhören, wenn eine Nichtübereinstimmung erkannt wird
Backport	HIVE-20098: Statistik: NPE beim Abrufen von Partitionsstatistiken für die Datumsspalte

Typ	Beschreibung
Backport	HIVE-20095 : Die Funktion zur Übertragung von Berechnungen auf externe JDBC-Tabellen wurde korrigiert
Backport	HIVE-20093 :: Verwendung mit Netty for Accounting LlapOutputFomatService ArrowBuf
Backport	HIVE-20090 : Erweitern Sie die Erstellung von Semijoin-Reduktionsfiltern, um neue Möglichkeiten zu entdecken
Backport	HIVE-20088 : Der Pfad für den Speicherort der Beeline-Konfiguration ist falsch zusammengestellt
Backport	HIVE-20082 : Bei der Konvertierung in eine Zeichenfolge wird HiveDecimal die Dezimalzahl nicht korrekt formatiert
Backport	HIVE-20069 : Korrektur der Reoptimierung bei Semijoin-Optimierung DPP
Backport	HIVE-20051 : Autorisierung für temporäre Tabellen überspringen
Backport	HIVE-20044 : Arrow Serde sollte Char-Werte korrekt auffüllen und leere Zeichenketten korrekt behandeln
Backport	HIVE-20028 : Die Metastore-Client-Cache-Konfiguration wird falsch verwendet
Backport	HIVE-20025 : Bereinigung von Ereignisdateien, die von erstellt wurden HiveProtoLoggingHook
Backport	HIVE-20020 : Hive contrib jar sollte nicht in lib sein

Typ	Beschreibung
Backport	HIVE-20013 : Fügt dem Datumstyp für die Funktion <code>to_date</code> eine implizite Umwandlung hinzu
Backport	HIVE-20011 : Verlassen Sie den Append-Modus im Proto-Logging-Hook
Backport	HIVE-20005 : <code>acid_table_stats</code> , <code>acid_no_buckets</code> usw. — Änderung der Abfrageergebnisse in der Filiale
Backport	HIVE-20004 : Die von 64 verwendete falsche Skala <code>ConvertDecimal</code> führt <code>ToDecimal</code> zu falschen Ergebnissen
Backport	HIVE-19995 : Gesamter Zeilenverkehr für ACID-Tabellen
Backport	HIVE-19993 : Die Verwendung eines Tabellenaalias, der auch als Spaltenname erscheint, ist nicht möglich
Backport	HIVE-19992 : Vektorisierung: Fortsetzung von -19951 --> Aufruf hinzufügen zu. <code>HIVE SchemaEvolution isOnlyImplicitKonvertierung</code> , um kodierte LLAP I/O nur dann zu deaktivieren, wenn die Datentypkonvertierung nicht implizit ORC ist
Backport	HIVE-19989 : Metastore verwendet einen falschen Anwendungsnamen für Metriken <code>HADOOP2</code>

Typ	Beschreibung
Backport	HIVE-19981 : Verwaltete Tabellen, die vom HiveStrictManagedMigration Hilfsprogramm in externe Tabellen konvertiert wurden, sollten so eingestellt werden, dass sie Daten löschen, wenn die Tabelle gelöscht wird
Backport	HIVE-19967 : SMB Join: Benötige Optraits für eine Operation PTFOperator GBY
Backport	HIVE-19935 : Hive-WM-Sitzung beendet: Die Anzahl der Aufgaben konnte nicht aktualisiert werden LLAP
Backport	HIVE-19924 : Distcp-Jobs kennzeichnen, die von Repl Load ausgeführt werden
Backport	HIVE-19891 : Das Einfügen in externe Tabellen mit benutzerdefinierten Partitionsverzeichnissen kann zu Datenverlust führen
Backport	HIVE-19850 : Das dynamische Bereinigen von Partitionen in Tez führt zu dem Fehler „Für den Tablescan wurde kein Job gefunden“
Backport	HIVE-19806 : Sortiert die QTests-Ausgabe, um Floskeln in den Testergebnissen zu vermeiden
Backport	HIVE-19770 : Support CBO für Abfragen mit mehreren gleichen Spalten in Select
Backport	HIVE-19769 : Erstellen Sie spezielle Objekte für Datenbank- und Tabellennamen
Backport	HIVE-19765 : Fügen Sie Parquet-spezifische Tests hinzu BlobstoreCliDriver

Typ	Beschreibung
Backport	HIVE-19759: Flockiger Test : # TestRpc testServerPort
Backport	HIVE-19711 : Das Hive-Schematool umgestalten
Backport	HIVE-19701 : getDelegationToken FromMetaStore muss nicht synchronisiert werden
Backport	HIVE-19694 : Die Create Materialized View-Anweisung sollte vor der Ausführung der MV-Anweisung nach MV-Namenskonflikten suchen. SQL
Backport	HIVE-19674 : Gruppieren nach Dezimalkonstanten wird in die Druid-Tabelle verschoben
Backport	HIVE-19668 : Über 30% des Heaps wurden durch doppelte org.antlr.runtime verschwendet. CommonToken's und doppelte Zeichenketten
Backport	HIVE-19663 : Die Generierung von IO-Berichten umgestalten LLAP
Backport	HIVE-19661 : Schalten Sie Hive UDFs so um, dass es die Re2J-Regex-Engine verwendet
Backport	HIVENPE-19628 : möglich in LLAP testSigning
Backport	HIVE-19568 : Aktiv/Passive HS2 HA: Direkte Verbindung zur passiven Instanz nicht zulassen HS2
Backport	HIVE-19564 : Vektorisierung: Probleme mit / Falschen Ergebnissen in der Arithmetik behalten NULL

Typ	Beschreibung
Backport	HIVE-19552 TestMiniDruidKafkaCliDriver: Aktiviere #druidkafkamini_basic .q
Backport	HIVE-19432 : GetTablesOperation ist zu langsam, wenn der Hive zu viele Datenbanken und Tabellen hat
Backport	HIVE-19360:CBO : Fügt dem Objekt ein „optimiertes“ Objekt hinzu SQL QueryPlan
Backport	HIVE-19326 : auto Sammeln von Statistiken: falsche Aggregation bei Abfragen UNION
Backport	HIVE-19313 : Tests schlagen fehl TestJdbcWith DBTokenStoreNoDoAs
Backport	HIVE-19285 : Fügt Logs zu den Unterklassen von hinzu MetaDataOperation
Backport	HIVE-19235 : Goldene Dateien für Minimr-Tests aktualisieren
Backport	HIVE-19104 : Wenn der Test mit Retry gestartet MetaStore wird, sollten die Instanzen unabhängig sein
Backport	HIVE-18986 : Beim Umbenennen von Tabellen wird java.lang ausgeführt. StackOverflowError in, dataNucleus wenn die Tabelle eine große Anzahl von Spalten enthält
Backport	HIVE-18920CBO : Initialisieren Sie die Janino-Anbieter vor der ersten Abfrage

Typ	Beschreibung
Backport	HIVE-18873 : Das unbeaufsichtigte Überspringen des Prädikats-Pushdowns für MR kann dazu führen, dass Speicherhandler ein falsches Ergebnis ausgeben HiveInputFormat
Backport	HIVE-18871 : Fehler bei der Ausführung von Hive on Tez, weil hive.aux.jars.path auf hdfs:// gesetzt wurde
Backport	HIVE-18725 : Verbessert die Fehlerbehandlung für Unterabfragen, wenn ein falscher Spaltenve rweis vorhanden ist
Backport	HIVE-18696 : Die Partitionsordner werden in der Methode .add_partitions_core möglicherweise nicht richtig bereinigt, wenn HiveMetaStore
Backport	HIVE-18453 :: Syntax "" hinzugefügt, um Unify & Parquet zu unterstützen ACID CREATE TRANSACTIONAL TABLE ACID ORC
Backport	HIVE-18201 : Deaktiviert _ für sq_count_chec XPROD EDGE
Backport	HIVE-18140 : Statistiken für partitionierte Tabellen können in gemischten Groß- und Kleinschreibung bei Basisstatistiken schief gehen
Backport	HIVE-17921 : Die Aggregation mit eingebund ener Struktur führt zu einem falschen Ergebnis LLAP
Backport	HIVE-17896 : TopNKey: Erzeugt einen eigenständigen vektorisierbaren T-Operator opNKey

Typ	Beschreibung
Backport	HIVE-17840 : frisst eine Ausnahme, wenn HiveMetaStore transactionalListeners notifyEvent scheitern
Backport	HIVE-17043 : Entferne nicht eindeutige Spalten aus der Gruppierung nach Schlüsseln, falls sie später nicht referenziert werden
Backport	HIVE-17040 : Eliminierung von Mitgliedern bei Vorhandensein einer FK-Beziehung
Backport	HIVE-16839 : Unsymmetrische Aufrufe von openTransaction/, commitTransaction wenn dieselbe Partition gleichzeitig geändert wird
Backport	HIVE-16100 : Der Optimierer für dynamisch sortierte Partitionen verliert gleichrangige Operatoren
Backport	HIVE-15956 : wenn viele Partitionen gelöscht werden StackOverflowError
Backport	HIVE-15177 : Die Authentifizierung mit Hive schlägt fehl, wenn der Kerberos-Authentifizierungstyp auf gesetzt ist und der Prinzipal _ enthält fromSubject HOST
Backport	HIVE-14898: Bei einem leeren Auth-Header-Fehler HS2 sollte der Callstack nicht protokolliert werden
Backport	HIVE-14493 : Partitionierungsunterstützung für materialisierte Ansichten
Backport	HIVE-14431 : COALESCE Erkennt als CASE

Typ	Beschreibung
Backport	HIVE-13457 : Endpunkte für die Überwachung von Informationen erstellen HS2 REST API
Backport	HIVE-12342 : Setzen Sie den Standardwert von <code>hive.optimize.index.filter</code> auf <code>true</code>
Backport	HIVE-10296 : Es wurde eine Cast-Ausnahme beobachtet, wenn Hive eine Multi-Join-Abfrage im Metastore ausführt
Backport	HIVE-6980 : Tabelle mithilfe von Direct SQL löschen

Amazon EMR 6.6.0 — Änderungen der Hive-Konfiguration

- Im Rahmen der OSS Änderung [HIVE-20703](#) wurde die Eigenschaft zum Sortieren dynamischer Partitionen, `hive.optimize.sort.dynamic.partition`, durch ersetzt. `hive.optimize.sort.dynamic.partition.threshold`

Die `hive.optimize.sort.dynamic.partition.threshold`-Konfiguration hat die folgenden möglichen Werte:

Wert	Beschreibung
0 (Standard)	Macht die Optimierung zur Sortierung dynamischer Partitionen zu einer kostenbasierten Entscheidung, wenn ORC Dateien verwendet werden. Die maximale Anzahl der in INSERT Abfragen zulässigen Schreibvorgänge wird berechnet auf der Grundlage von (Executor-/Container-Speicher) * (Prozentsatz des von orc belegten Speichers) geteilt durch den maximalen Speicher (Stripe-Größe), der von einem einzelnen Writer belegt wird.
-1	Deaktiviert die Optimierung zur vollständigen Sortierung dynamischer Partitionen.

Wert	Beschreibung
1	Aktiviert die globale Sortierung dynamischer Partitionen. Dadurch bleibt für jeden Partitionswert im Reducer nur ein Record Writer geöffnet, wodurch der Speicherdruck auf den Reducern reduziert wird.
2 (oder eine größere Ganzzahl)	Weist Hive an, die angegebene Ganzzahl als Schwellenwert für die maximale Anzahl von Autoren zu verwenden.

Amazon EMR 6.6.0 — Bekannte Probleme mit Hive

- Abfragen mit Fensterfunktionen für dieselbe Spalte wie Join können zu ungültigen Transformationen führen, wie in [HIVE-25278](#) gemeldet, und zu falschen Ergebnissen oder Abfragefehlern führen. Um das Problem zu umgehen, können Sie solche Abfragen auf CBO Abfrageebene deaktivieren. Wenden Sie sich für weitere Informationen an den AWS Support.
- Amazon EMR 6.6.0 enthält die Hive-Softwareversion 3.1.2. [Hive 3.1.2 führt eine Funktion ein, die Textdateien aufteilt, wenn sie eine Kopf- und Fußzeile enthalten \(-21924\). HIVE](#) Der Apache Tez App Master liest jede Ihrer Dateien, um Versatzpunkte im Datenbereich zu bestimmen. Diese Verhaltensweisen zusammen können sich negativ auf die Leistung auswirken, wenn Ihre Abfragen eine große Anzahl kleiner Textdateien lesen. Um dieses Problem zu umgehen, können Sie `CombineHiveInputFormat` und die maximale Teilungsgröße verwenden und anpassen, indem Sie die folgenden Eigenschaften konfigurieren:

```
SET hive.tez.input.format=org.apache.hadoop.hive ql.io.CombineHiveInputFormat;
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT Abfragen mit dynamischer Partition und einer ORDER BY- oder SORT BY-Klausel immer zwei Reducer. Dieses Problem wird durch die OSS Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde in OSS Hive als Teil von [HIVE-22269](#) und in [Amazon](#) 6.10.0 behoben. EMR

Hudi

[Apache Hudi](#) ist ein Open-Source-Datenverwaltungs-Framework, das verwendet wird, um die inkrementelle Datenverarbeitung und die Entwicklung von Datenpipelines zu vereinfachen, indem Einfüge-, Aktualisierungs-, Upsert- und Löschfunktionen auf Datensatzebene bereitgestellt werden. Upsert bezieht sich auf die Möglichkeit, Datensätze in ein vorhandenes Dataset einzufügen, wenn sie noch nicht vorhanden sind, oder sie zu aktualisieren, falls sie schon vorhanden sind. Durch die effiziente Verwaltung der Datenaufstellung in Amazon S3 ermöglicht Hudi die Aufnahme und Aktualisierung von Daten in nahezu Echtzeit. Hudi verwaltet sorgfältig Metadaten der Aktionen, die für das Dataset ausgeführt werden, um leichter sicherzustellen zu können, dass die Aktionen atomar und konsistent sind.

Hudi ist in [Apache Spark](#), [Apache Hive](#) und [Presto](#) integriert. In den EMR Amazon-Release-Versionen 6.1.0 und höher ist Hudi auch in [Trino](#) (Presto) integriert. SQL

EMR installiert EMR ab Amazon-Version 5.28.0 standardmäßig Hudi-Komponenten, wenn Spark, Hive, Presto oder Flink installiert sind. Sie können Spark oder das DeltaStreamer Hudi-Hilfsprogramm verwenden, um Hudi-Datensätze zu erstellen oder zu aktualisieren. Sie können Hive, Spark, Presto oder Flink verwenden, um ein Hudi-Datensatz interaktiv abzufragen oder Datenverarbeitungspipelines mithilfe des inkrementellen Pull-Vorgangs zu erstellen. Inkrementeller Pull-Vorgang bezieht sich auf die Fähigkeit, nur die Daten zu extrahieren, die sich zwischen zwei Aktionen geändert haben.

Aufgrund dieser Features eignet sich Hudi für die folgenden Anwendungsfälle:

- Arbeiten mit Streaming-Daten von Sensoren und anderen IoT-Geräten (Internet of Things), die bestimmte Dateneinfüge- und Aktualisierungsereignisse erfordern.
- Einhaltung der Datenschutzvorschriften in Anwendungen, in denen Benutzer möglicherweise vergessen werden möchten oder ihre Einwilligung für die Verwendung ihrer Daten ändern möchten.
- Implementierung eines [Systems zur Erfassung von Änderungsdaten \(CDC\)](#), mit dem Sie im Laufe der Zeit Änderungen an einem Datensatz vornehmen können.

In der folgenden Tabelle sind die Version von Hudi aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Hudi EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hudi installiert wurden, finden Sie unter Komponentenversionen von Version 7.2.0.](#)

Hudi-Versionsinformationen für emr-7.2.0


EMRAmazon-Freigabeetikett	Hudi-Version	Mit Hudi installierte Komponenten
emr-7.2.0	Hudi 0.14.1-amzn-1	Not available.

In der folgenden Tabelle sind die Version von Hudi aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Hudi EMR installiert.

Die Version der Komponenten, die mit Hudi in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0.](#)

Hudi-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Hudi-Version	Mit Hudi installierte Komponenten
emr-6.15.0	Hudi 0.14.0-amzn-0	Not available.

 Note

Die EMR Amazon-Version 6.8.0 enthält [Apache Hudi](#) 0.11.1. Amazon EMR 6.8.0-Cluster sind jedoch auch mit der Open-Source-Version von Hudi 0.12.0 kompatibel. `hudi-spark3.3-bundle_2.12`

In der folgenden Tabelle sind die Version von Hudi aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Hudi EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hudi installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Hudi-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Hudi-Version	Mit Hudi installierte Komponenten
emr-5.36.2	Hudi 0.10.1-amzn-1	Not available.

Themen

- [Funktionsweise von Hudi](#)
- [Überlegungen und Einschränkungen für die Verwendung von Hudi bei Amazon EMR](#)
- [Erstellen eines Cluster mit installierter Hudi-Anwendung](#)
- [Arbeiten mit einem Hudi-Datensatz](#)
- [Benutze den Hudi CLI](#)
- [Hudi-Versionsverlauf](#)

Funktionsweise von Hudi

Wenn Sie Hudi mit Amazon verwenden EMR, können Sie mit der Spark-Datenquelle API oder dem DeltaStreamer Hudi-Hilfsprogramm Daten in den Datensatz schreiben. Hudi organisiert einen Datensatz in einer partitionierten Verzeichnisstruktur unter einer *basepath*, die einer herkömmlichen Hive-Tabelle ähnelt. Die Besonderheiten, wie die Daten in diesen Verzeichnissen als Dateien angelegt werden, hängen von dem von Ihnen ausgewählten Dataset-Typ ab. Sie können entweder Copy on Write (CoW, Kopieren beim Schreiben) oder Merge on Read (MoR, Beim Lesen zusammenführen) wählen.

Unabhängig vom Dataset-Typ wird jede Partition in einem Dataset eindeutig durch ihren *partitionpath* relativ zum *basepath* identifiziert. Innerhalb jeder Partition werden Datensätze in mehrere Datendateien verteilt. Weitere Informationen finden Sie unter [Dateimanagement](#) in der Apache Hudi-Dokumentation.

Jede Aktion in Hudi hat einen entsprechenden Commit, der durch einen monoton ansteigenden Zeitstempel gekennzeichnet ist, der als Instant bezeichnet wird. Hudi speichert eine Reihe aller am Datensatz ausgeführten Aktionen als Zeitleiste. Hudi nutzt die Timeline, um eine Snapshot-Isolation zwischen Lesern und Autoren zu ermöglichen und einen Rollback zu einem früheren Zeitpunkt zu ermöglichen. Weitere Informationen zu den von Hudi aufgezeichneten Aktionen und dem Status der Aktionen finden Sie unter [Timeline](#) in der Apache-Hudi-Dokumentation.

Grundlegendes zu den Speichertypen von Datensätzen: „Beim Schreiben kopieren“ gegenüber „beim Lesen zusammenführen“

Wenn Sie ein Hudi-Datensatz erstellen, geben Sie an, dass das Datensatz entweder beim Schreiben kopiert oder beim Lesen zusammengeführt werden soll.

- Copy on Write (CoW, Beim Schreiben kopieren) – Daten werden in einem spaltenbasierten Format (Parquet) gespeichert, und jedes Update erstellt während eines Schreibvorgangs eine neue Version von Dateien. CoW ist der Standard Speichertyp.
- Merge on Read (MoR, Beim Lesen zusammenführen) – Daten werden mit einer Kombination aus spalten- (Parquet) und zeilenbasierten (Avro) Formaten gespeichert. Updates werden in zeilenbasierten Delta-Dateien protokolliert und nach Bedarf komprimiert, um neue Versionen der Spaltendateien zu erstellen.

Bei CoW-Datasets wird jedes Mal, wenn ein Datensatz aktualisiert wird, die Datei, die den Datensatz enthält, mit den aktualisierten Werten neu geschrieben. Bei einem MoR-Datensatz schreibt Hudi jedes Mal, wenn es eine Aktualisierung gibt, nur die Zeile für den geänderten Datensatz. MoR eignet sich besser für schreib- oder änderungsintensive Workloads mit weniger Lesevorgängen. CoW eignet sich besser für leseintensive Workloads für Daten, die sich seltener ändern.

Hudi bietet drei logische Ansichten für den Zugriff auf die Daten:

- Lese-optimierte Ansicht – Stellt den neuesten festgeschriebenen Datensatz aus CoW-Tabellen und den neuesten komprimierten Datensatz aus MoR-Tabellen bereit.
- Inkrementelle Ansicht — Stellt einen Änderungsstream zwischen zwei Aktionen aus einem CoW-Datensatz bereit, um Downstream-Jobs zuzuführen und Workflows zu extrahieren, zu transformieren und zu laden (ETL).
- Echtzeitansicht – Stellt die neuesten festgeschriebenen Daten aus einer MoR-Tabelle bereit, indem die spalten- und zeilenbasierten Dateien inline zusammengeführt werden.

Wenn Sie die leseoptimierte Ansicht abfragen, gibt die Abfrage alle komprimierten Daten zurück, enthält jedoch nicht die neuesten Delta-Commits. Das Abfragen dieser Daten bietet eine gute Leseleistung, aber die neuesten Daten werden nicht mehr angezeigt. Wenn Sie die Echtzeitansicht abfragen, führt Hudi die komprimierten Daten beim Lesen mit den Delta-Commits zusammen. Die neuesten Daten stehen für die Abfrage zur Verfügung, aber der Rechenaufwand für die Zusammenführung macht die Abfrage weniger performant. Die Möglichkeit, komprimierte Daten oder

Echtzeitdaten abzufragen, ermöglicht es Ihnen, bei der Abfrage zwischen Leistung und Flexibilität zu wählen.

Weitere Informationen zu den Kompromissen zwischen Speichertypen finden Sie unter [Speichertypen und Ansichten](#) in der Apache-Hudi-Dokumentation.

Hudi erstellt zwei Tabellen im Hive-Metastore für MoR: eine Tabelle mit dem von Ihnen angegebenen Namen, bei der es sich um eine leseoptimierte Ansicht handelt, und eine Tabelle mit demselben Namen, an den `_rt` angehängt wird, bei der es sich um eine Echtzeitansicht handelt. Sie können beide Tabellen abfragen.

Registrieren eines Hudi-Datensatzes bei Ihrem Metastore

Wenn Sie eine Hudi-Tabelle im Hive-Metastore registrieren, können Sie Hudi-Tabellen mit Hive, Spark SQL oder Presto wie mit jeder anderen Tabelle abfragen. Darüber hinaus können Sie Hudi in AWS Glue integrieren, indem Sie Hive und Spark so konfigurieren, dass sie den AWS Glue-Datenkatalog als Metastore verwenden. Hudi registriert für MoR-Tabellen den Datensatz als zwei Tabellen im Metastore: eine Tabelle mit dem von Ihnen angegebenen Namen, bei der es sich um eine leseoptimierte Ansicht handelt, und eine Tabelle mit demselben Namen, an die `_rt` angehängt wird, bei dem es sich um eine Echtzeitansicht handelt.

Sie registrieren eine Hudi-Tabelle mit dem Hive-Metastore, wenn Sie mit Spark ein Hudi-Datensatz erstellen, indem Sie die Option `HIVE_SYNC_ENABLED_OPT_KEY` auf `"true"` einstellen und andere erforderliche Eigenschaften bereitstellen. Weitere Informationen finden Sie unter [Arbeiten mit einem Hudi-Datensatz](#). Darüber hinaus können Sie das Befehlszeilenprogramm `hive_sync_tool` verwenden, um einen Hudi-Datensatz separat als Tabelle in Ihrem Metastore zu registrieren.

Überlegungen und Einschränkungen für die Verwendung von Hudi bei Amazon EMR

- Datensatzschlüselfeld darf nicht null oder leer sein – Das Feld, das Sie als Datensatzschlüselfeld angeben, darf weder `null`-Werte noch leere Werte aufweisen.
- Das Schema wird standardmäßig beim Upsert und Insert aktualisiert — Hudi bietet eine Schnittstelle, `HoodieRecordPayload` die festlegt, wie die Eingabe `DataFrame` und der vorhandene Hudi-Datensatz zusammengeführt werden, um einen neuen, aktualisierten Datensatz zu erzeugen. Hudi bietet eine Standardimplementierung dieser Klasse `OverwriteWithLatestAvroPayload`, die bestehende Datensätze überschreibt

und das in der Eingabe angegebene Schema aktualisiert. DataFrame Um diese Logik für die Implementierung von Merge- und Teilaktualisierungen anzupassen, können Sie mithilfe des Parameters `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY` eine Implementierung der `HoodieRecordPayload`-Schnittstelle bereitstellen.

- Löschen erfordert Schema – Beim Löschen müssen Sie den Datensatzschlüssel, den Partitionsschlüssel und die Schlüsselfelder für die Vorab-Kombination angeben. Andere Spalten können auf `null` gesetzt werden oder leer sein, aber das vollständige Schema ist erforderlich.
- Einschränkungen bei MoR-Tabellen – MoR-Tabellen unterstützen kein Savepointing. Sie können MoR-Tabellen mit der leseoptimierten Ansicht oder der Echtzeitansicht (`tableName_rt`) von SparkSQL, Presto oder Hive abfragen. Wenn Sie die leseoptimierte Ansicht verwenden, werden nur Basisdateidaten und keine zusammengeführte Ansicht von Basis- und Protokolldaten angezeigt.
- Hive
 - Für die Registrierung von Tabellen im Hive-Metastore erwartet, dass der Hive Thrift-Server am Standardport `10000` ausgeführt wird. Wenn Sie diesen Port mit einem benutzerdefinierten Port überschreiben, übergeben Sie die Option `HIVE_URL_OPT_KEY` wie im folgenden Beispiel gezeigt.

```
.option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number
```

- Der Datentyp `timestamp` in Spark ist in Hive als `long`-Datentyp und nicht als Hive-Typ `timestamp` registriert.
 - Presto
 - Presto unterstützt nicht das Lesen von MoR-Echtzeitabfragen. Funktionen in Hudi-Versionen unter `0.6.0`.
 - Presto unterstützt nur Snapshot-Abfragen.
 - Damit Presto Hudi-Datensatz-Spalten korrekt interpretiert, stellen Sie den `hive.parquet_use_column_names`-Wert auf `true` ein.
 - Um den Wert für eine Sitzung festzulegen, führen Sie in der Presto-Shell den folgenden Befehl aus:
- ```
set session hive.parquet_use_column_names=true
```
- Um den Wert auf Clusterebene festzulegen, verwenden Sie die Konfigurationsklassifizierung `presto-connector-hive`, um `hive.parquet.use_column_names` auf `true`

einzustellen, wie im folgenden Beispiel veranschaulicht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
[
 {
 "Classification": "presto-connector-hive",
 "Properties": {
 "hive.parquet.use-column-names": "true"
 }
 }
]
```

- HBaseIndex
  - Die HBase Version, die zum Erstellen von Hudi verwendet wurde, unterscheidet sich möglicherweise von der Version, die im EMR Versionshandbuch aufgeführt ist. Um die richtigen Abhängigkeiten für Ihre Spark-Sitzung abzurufen, führen Sie den folgenden Befehl aus.

```
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"
```

## Erstellen eines Cluster mit installierter Hudi-Anwendung

Mit der EMR Amazon-Release-Version 5.28.0 und höher EMR installiert Amazon standardmäßig Hudi-Komponenten, wenn Spark, Hive oder Presto installiert sind. Um Hudi auf Amazon zu verwenden EMR, erstellen Sie einen Cluster mit einer oder mehreren der folgenden installierten Anwendungen:

- Hadoop
- Hive
- Spark
- Presto
- Flink

Sie können einen Cluster mit dem AWS Management Console AWS CLI, dem oder dem Amazon erstellen EMR API.

## Um einen Cluster mit Hudi zu erstellen, verwenden Sie AWS Management Console

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie unter „Software Configuration (Softwarekonfiguration)“ emr-5.28.0 oder höher für Release aus und wählen Sie Hadoop, Hive, Spark, Presto und Tez zusammen mit anderen Anwendungen aus, die Ihr Cluster benötigt.
4. Konfigurieren Sie nach Bedarf andere Optionen für Ihre Anwendung und wählen Sie dann Next (Weiter).
5. Konfigurieren Sie die Optionen für Hardware und General cluster settings (Allgemeine Clustereinstellungen) ganz nach Wunsch.
6. Für Sicherheitsoptionen empfehlen wir, dass Sie ein EC2key pair auswählen, mit dem Sie eine Verbindung zur Befehlszeile des Master-Knotens herstellen könnenSSH. Auf diese Weise können Sie die in diesem Handbuch beschriebenen CLI Spark-Shell-Befehle, Hive-Befehle und CLI Hudi-Befehle ausführen.
7. Wählen Sie weitere Sicherheitsoptionen wie gewünscht und wählen Sie anschließend Create cluster (Cluster erstellen) aus.

## Arbeiten mit einem Hudi-Datensatz

Hudi unterstützt das Einfügen, Aktualisieren und Löschen von Daten in Hudi-Datensätzen über Spark. Weitere Informationen finden Sie unter [Hudi-Tabellen schreiben](#) in der Apache-Hudi-Dokumentation.

Die folgenden Beispiele zeigen, wie Sie die interaktive Spark-Shell starten, Spark Submit verwenden oder Amazon EMR Notebooks verwenden, um mit Hudi auf Amazon EMR zu arbeiten. Sie können auch das DeltaStreamer Hudi-Hilfsprogramm oder andere Tools verwenden, um in einen Datensatz zu schreiben. In diesem Abschnitt zeigen die Beispiele, wie Sie mit Datensätzen arbeiten, indem Sie die Spark-Shell verwenden, während Sie SSH als hadoop Standardbenutzer mit dem Master-Knoten verbunden sind.

## Starten Sie die Spark-Shell mit Amazon EMR 6.7 und höher

Wenn Sie Amazon EMR 6.7.0 oder höher ausführen `spark-shell` oder `spark-sql` verwenden, übergeben Sie die folgenden Befehle. `spark-submit`

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

So öffnen Sie die Spark-Shell auf dem Primärknoten

1. Stellen Sie mithilfe von Connect zum Primärknoten herSSH. Weitere Informationen finden Sie unter [Connect mit dem primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. Um die PySpark Shell zu verwenden, ersetzen Sie `spark-shell` mit `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"
```

## Starten Sie die Spark-Shell mit Amazon EMR 6.6 und früher

Wenn Sie Amazon EMR 6.6.x oder früher ausführen `spark-shell` oder `spark-sql` verwenden, übergeben Sie die folgenden Befehle. `spark-submit`

### Note

- Amazon EMR 6.2 und 5.31 und höher (Hudi 0.6.x und höher) können das in der Konfiguration weglassen. `spark-avro.jar`
- Amazon EMR 6.5 und 5.35 und höher (Hudi 0.9.x und höher) können in der Konfiguration weglassen `spark.sql.hive.convertMetastoreParquet=false`.

- Amazon EMR 6.6 und 5.36 und höher (Hudi 0.10.x und höher) müssen die HoodieSparkSessionExtension Konfiguration enthalten, wie sie im Spark-Leitfaden für [Version: 0.10.0 beschrieben ist](#):

```
--conf
"spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
\
```

So öffnen Sie die Spark-Shell auf dem Primärknoten

1. Stellen Sie mithilfe von Connect zum Primärknoten her SSH. Weitere Informationen finden Sie unter [Connect mit dem primären Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. Um die PySpark Shell zu verwenden, ersetzen Sie *spark-shell* mit *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
```

## Verwenden Sie Hudi mit Amazon EMR Notebooks mit Amazon EMR 6.7 und höher

Um Hudi mit Amazon EMR Notebooks zu verwenden, müssen Sie zuerst die Hudi-JAR-Dateien aus dem lokalen Dateisystem HDFS auf den Master-Knoten des Notebook-Clusters kopieren. Anschließend verwenden Sie den Notebook-Editor, um Ihr EMR Notebook für die Verwendung von Hudi zu konfigurieren.

Um Hudi mit Amazon EMR Notebooks zu verwenden

1. Erstellen und starten Sie einen Cluster für Amazon EMR Notebooks. Weitere Informationen finden Sie unter [Erstellen von EMR Amazon-Clustern für Notebooks](#) im Amazon EMR Management Guide.
2. Stellen Sie mithilfe der JAR-Dateien eine Connect zum Master-Knoten des Clusters her SSH und kopieren Sie sie dann aus dem lokalen Dateisystem in, HDFS wie in den folgenden

Beispielen gezeigt. In diesem Beispiel erstellen wir aus HDFS Gründen der Übersichtlichkeit der Dateiverwaltung ein Verzeichnis. Falls gewünscht HDFS, können Sie Ihr eigenes Ziel wählen.

```
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. Öffnen Sie den Notebook-Editor, geben Sie den Code aus dem folgenden Beispiel ein und führen Sie ihn aus.

```
%%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"
}}
```

## Verwenden Sie Hudi mit Amazon EMR Notebooks mit Amazon EMR 6.6 und früher

Um Hudi mit Amazon EMR Notebooks zu verwenden, müssen Sie zuerst die Hudi-JAR-Dateien aus dem lokalen Dateisystem HDFS auf den Master-Knoten des Notebook-Clusters kopieren. Anschließend verwenden Sie den Notebook-Editor, um Ihr EMR Notebook für die Verwendung von Hudi zu konfigurieren.

Um Hudi mit Amazon EMR Notebooks zu verwenden

1. Erstellen und starten Sie einen Cluster für Amazon EMR Notebooks. Weitere Informationen finden Sie unter [Erstellen von EMR Amazon-Clustern für Notebooks](#) im Amazon EMR Management Guide.
2. Stellen Sie mithilfe der JAR-Dateien eine Connect zum Master-Knoten des Clusters her SSH und kopieren Sie sie dann aus dem lokalen Dateisystem in, HDFS wie in den folgenden Beispielen gezeigt. In diesem Beispiel erstellen wir aus HDFS Gründen der Übersichtlichkeit der Dateiverwaltung ein Verzeichnis. Falls gewünscht HDFS, können Sie Ihr eigenes Ziel wählen.

```
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. Öffnen Sie den Notebook-Editor, geben Sie den Code aus dem folgenden Beispiel ein und führen Sie ihn aus.

```
{ "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"
}}
```

## Initialisieren Sie eine Spark-Sitzung für Hudi

Wenn Sie Scala verwenden, müssen Sie die folgenden Klassen in Ihre Spark-Sitzung importieren. Dies muss einmal pro Spark-Sitzung erfolgen.

```
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
```

## Schreiben Sie in einen Hudi-Datensatz

Die folgenden Beispiele zeigen, wie Sie einen Hudi-Datensatz erstellen DataFrame und ihn als Hudi-Datensatz schreiben.

**Note**

Um Codebeispiele in die Spark-Shell einzufügen, geben Sie an der Eingabeaufforderung **:paste** ein, fügen das Beispiel ein und drücken dann **CTRL + D**.

Jedes Mal, wenn Sie einen DataFrame in einen Hudi-Datensatz schreiben, müssen Sie Folgendes angeben. `DataSourceWriteOptions` Viele dieser Optionen sind unter den Schreiboperationen wahrscheinlich identisch. Im folgenden Beispiel werden allgemeine Optionen unter Verwendung der Variablen `hudiOptions` angegeben, die von nachfolgenden Beispielen verwendet wird.

Schreiben Sie mit Scala mit Amazon EMR 6.7 und höher

**Note**

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
// 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.MultiPartKeyValueExtractor",
HoodieSyncConfig.META_SYNC_ENABLED.key -> "true",
HiveSyncConfig.HIVE_SYNC_MODE.key -> "hms",
HoodieSyncConfig.META_SYNC_TABLE_NAME.key -> "tableName",
HoodieSyncConfig.META_SYNC_PARTITION_FIELDS.key -> "creation_date"
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
 .format("hudi")
 .options(hudiOptions)
 .option(DataSourceWriteOptions.OPERATION_OPT_KEY,"insert")
 .mode(SaveMode.Overwrite)
 .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))

```

Schreiben Sie mit Scala mit Amazon EMR 6.6 und früher

```

// 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[MultiPartKeyValueExtractor].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/"))
```

## Schreiben Sie mit PySpark

```
Create a DataFrame
inputDF = spark.createDataFrame(
 [
 ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
 ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
 ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
 ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
 ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
 ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z"),
],
 ["id", "creation_date", "last_update_time"]
)

Specify common DataSourceWriteOptions in the single hudiOptions variable
hudiOptions = {
'hoodie.table.name': 'tableName',
'hoodie.datasource.write.recordkey.field': 'id',
'hoodie.datasource.write.partitionpath.field': 'creation_date',
'hoodie.datasource.write.precombine.field': 'last_update_time',
'hoodie.datasource.hive_sync.enable': 'true',
'hoodie.datasource.hive_sync.table': 'tableName',
'hoodie.datasource.hive_sync.partition_fields': 'creation_date',
'hoodie.datasource.hive_sync.partition_extractor_class':
'org.apache.hudi.hive.MultiPartKeysValueExtractor'
}

Write a DataFrame as a Hudi dataset
inputDF.write \
 .format('org.apache.hudi') \
 .option('hoodie.datasource.write.operation', 'insert') \
 .options(**hudiOptions) \
 .mode('overwrite') \
```

```
.save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

### Note

Möglicherweise sehen Sie in Codebeispielen und Benachrichtigungen „hoodie“ anstelle von Hudi. In der Hudi-Codebasis wird häufig die alte Schreibweise „hoodie“ verwendet.

## DataSourceWriteOptions Referenz für Hudi

| Option                      | Beschreibung                                                                                                                                                                                                                                                                                                                         |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| TABLE_NAME                  | Der Tabellename, unter dem das Dataset registriert werden soll.                                                                                                                                                                                                                                                                      |
| TABLE_TYPE_OPT_KEY          | Optional. Gibt an, ob das Dataset als "COPY_ON_WRITE" oder "MERGE_ON_READ" erstellt wird. Der Standardwert ist "COPY_ON_WRITE".                                                                                                                                                                                                      |
| RECORDKEY_FIELD_OPT_KEY     | Das Datensatzschlüselfeld, dessen Wert als die <code>recordKey</code> -Komponente von <code>HoodieKey</code> verwendet wird. Der tatsächliche Wert wird durch Aufrufen von <code>.toString()</code> für den Feldwert abgerufen. Verschachtelte Felder können mithilfe der Punktnotation angegeben werden, z. B. <code>a.b.c</code> . |
| PARTITIONPATH_FIELD_OPT_KEY | Das Partitionspfadfeld, dessen Wert als die Komponente <code>partitionPath</code> von <code>HoodieKey</code> verwendet wird. Der tatsächliche Wert wird durch Aufrufen von <code>.toString()</code> für den Feldwert abgerufen.                                                                                                      |
| PRECOMBINE_FIELD_OPT_KEY    | Das Feld, das in der Vorab-Kombination vor dem tatsächlichen Schreiben verwendet wird. Wenn zwei Datensätze denselben Schlüsselwert haben, wählt Hudi den Datensatz mit dem größten Wert für das Vorab-Kombinationsfeld.                                                                                                             |

| Option | Beschreibung                                              |
|--------|-----------------------------------------------------------|
|        | wie von <code>Object.compareTo(..)</code> festgelegt aus. |

Die folgenden Optionen sind nur erforderlich, um die Hudi-Datensatz-Tabelle in Ihrem Metastore zu registrieren. Wenn Sie Ihr Hudi-Datensatz nicht als Tabelle im Hive-Metastore registrieren, sind diese Optionen nicht erforderlich.

#### DataSourceWriteOptions Referenz für Hive

| Option                                 | Beschreibung                                                                                                                         |
|----------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| HIVE_DATABASE_OPT_KEY                  | Die Hive-Datenbank, mit der synchronisiert werden soll. Der Standardwert ist "default".                                              |
| HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY | Die Klasse, mit der Partitionsfeldwerte in Hive-Partitionsspalten extrahiert werden.                                                 |
| HIVE_PARTITION_FIELDS_OPT_KEY          | Das Feld im Dataset, anhand dessen Hive-Partitionsspalten bestimmt werden sollen.                                                    |
| HIVE_SYNC_ENABLED_OPT_KEY              | Wenn diese Option auf "true" eingestellt ist, wird das Dataset beim Apache-Hive-Metastore registriert. Der Standardwert ist "false". |
| HIVE_TABLE_OPT_KEY                     | Erforderlich Der Name der Tabelle in Hive, mit der synchronisiert werden soll. z. B. "my_hudi_table_cow".                            |
| HIVE_USER_OPT_KEY                      | Optional. Der Hive-Benutzername, der bei der Synchronisierung verwendet werden soll. z. B. "hadoop".                                 |
| HIVE_PASS_OPT_KEY                      | Optional. Das von HIVE_USER_OPT_KEY angegebene Hive-Passwort für den Benutzer.                                                       |
| HIVE_URL_OPT_KEY                       | Der Hive-Metastore. URL                                                                                                              |

## Upsert Daten

Das folgende Beispiel zeigt, wie Daten durch Schreiben von A verändert werden. DataFrame  
Im Gegensatz zum vorherigen Einfügebeispiel wird der Wert OPERATION\_OPT\_KEY auf UPSERT\_OPERATION\_OPT\_VAL eingestellt. Darüber hinaus wird mit `.mode(SaveMode.Append)` angegeben, dass der Datensatz angehängt werden soll.

Upsert mit Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
// 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/myhudidataset/"))
```

Upsert mit Scala mit Amazon EMR 6.6 und früher

```
// Create a new DataFrame from the first row of inputDF with a different creation_date value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
 .format("org.apache.hudi")
 .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
 DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
 .options(hudiOptions))
```

```
.mode(SaveMode.Append)
.save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/")
```

## Upsert mit 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/')
```

## Einen Datensatz löschen

Um einen Datensatz dauerhaft zu löschen, können Sie eine leere Datenlast einfügen. In diesem Fall gibt die Option `PAYLOAD_CLASS_OPT_KEY` die Klasse `EmptyHoodieRecordPayload` an. Im Beispiel wird derselbe DataFrame, verwendet `updateDF`, der im Upsert-Beispiel verwendet wurde, um denselben Datensatz anzugeben.

Löschen mithilfe von Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
(updateDF.write
 .format("hudi")
 .options(hudiOptions)
 .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete"))
```

```
.mode(SaveMode.Append)
.save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

## Löschen mithilfe von Scala mit Amazon EMR 6.6 und früher

```
(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/"))
```

## Löschen mit 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/')
```

Sie können Daten auch dauerhaft löschen, indem Sie `OPERATION_OPT_KEY` auf `DELETE_OPERATION_OPT_VAL` einstellen, um alle Datensätze in dem von Ihnen eingereichten Datensatz zu entfernen. Anweisungen zur Durchführung von „weichen Löschungen“ und weitere Informationen zum Löschen von Daten, die in Hudi-Tabellen gespeichert sind, finden Sie unter [Löschen](#) in der Apache-Hudi-Dokumentation.

## Aus einem Hudi-Datensatz lesen

Um Daten zum aktuellen Zeitpunkt abzurufen, führt Hudi standardmäßig Snapshot-Abfragen durch. Im Folgenden finden Sie ein Beispiel für die Abfrage des in S3 geschriebenen Datensatzes in [Schreiben Sie in einen Hudi-Datensatz](#). Ersetzen `s3://DOC-EXAMPLE-BUCKET/myhuidataset` mit Ihrem Tabellenpfad und fügen Sie Platzhalter-Sternchen für jede Partitionsebene sowie ein zusätzliches Sternchen hinzu. In diesem Beispiel gibt es eine Partitionsebene, daher haben wir zwei Platzhaltersymbole hinzugefügt.

## Lesen Sie mit Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi](#) 0.11.0-amzn-0, das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
val snapshotQueryDF = spark.read
 .format("hudi")
 .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset")
 .show()
```

## Lesen Sie mit Scala mit Amazon EMR 6.6 und früher

```
(val snapshotQueryDF = spark.read
 .format("org.apache.hudi")
 .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" + "/*/*"))

snapshotQueryDF.show()
```

## Lesen Sie mit PySpark

```
snapshotQueryDF = spark.read \
 .format('org.apache.hudi') \
 .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset' + '/*/*')

snapshotQueryDF.show()
```

## Inkrementelle Abfragen

Sie können mit Hudi auch inkrementelle Abfragen durchführen, um einen Stream von Datensätzen abzurufen, die sich seit einem bestimmten Commit-Zeitstempel geändert haben. Setzen Sie dazu das Feld `QUERY_TYPE_OPT_KEY` auf `QUERY_TYPE_INCREMENTAL_OPT_VAL`. Fügen Sie dann einen Wert für `BEGIN_INSTANTTIME_OPT_KEY` hinzu, um alle Datensätze abzurufen, die seit dem angegebenen Zeitpunkt geschrieben wurden. Inkrementelle Abfragen sind in der Regel zehnmal effizienter als ihre Gegenstücke im Batch-Modus, da sie nur geänderte Datensätze verarbeiten.



Wenn Sie inkrementelle Abfragen ausführen, verwenden Sie den Pfad der Stammtabelle (Basistabelle) ohne die für Snapshot-Abfragen verwendeten Platzhaltersterchen.

### Note

Presto unterstützt keine inkrementellen Abfragen.

## Inkrementelle Abfragen mit 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()
```

## Inkrementelle Abfragen mit PySpark

```
readOptions = {
 'hoodie.datasource.query.type': 'incremental',
 'hoodie.datasource.read.begin.instanttime': <beginInstantTime>,
}

incQueryDF = spark.read \
 .format('org.apache.hudi') \
 .options(**readOptions) \
 .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset')

incQueryDF.show()
```

Weitere Informationen zum Lesen von Hudi-Datensätzen finden Sie unter [Abfragen von Hudi-Tabellen](#) in der Apache-Hudi-Dokumentation.

## Benutze den Hudi CLI

Sie können das Hudi verwendenCLI, um Hudi-Datensätze zu verwalten, um Informationen über Commits, das Dateisystem, Statistiken und mehr anzuzeigen. Sie können den auch verwenden, um Verdichtungen manuell durchzuführenCLI, Komprimierungen zu planen oder geplante Verdichtungen

abzubrechen. Weitere Informationen finden Sie unter [Interacting via CLI](#) in der Apache Hudi-Dokumentation.

Um Hudi zu starten CLI und eine Verbindung zu einem Datensatz herzustellen

1. Connect zum Master-Knoten her mit SSH. Weitere Informationen finden Sie unter [Connect dem Master-Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie in die Befehlszeile `/usr/lib/hudi/cli/bin/hudi-cli.sh` ein.

Die Eingabeaufforderung ändert sich in `hudi->`.

3. Geben Sie den folgenden Code ein, um eine Verbindung mit einem Dataset herzustellen. Ersetzen `s3://DOC-EXAMPLE-BUCKET/myhuidataset` mit dem Pfad zu dem Datensatz, mit dem Sie arbeiten möchten. Der von uns verwendete Wert entspricht dem Wert, der in früheren Beispielen festgelegt wurde.

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhuidataset
```

Die Eingabeaufforderung ändert sich, um das Dataset einzuschließen, mit dem Sie verbunden sind, wie im folgenden Beispiel gezeigt.

```
hudi:myhuidataset->
```

## Hudi-Versionsverlauf

In der folgenden Tabelle sind die Versionen von Hudi aufgeführt, die in jeder Release-Version von Amazon enthalten sind EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Versionsinformationen von Hudi

| Amazon EMR Release-Etikett | Hudi-Version  | Mit Hudi installierte Komponenten |
|----------------------------|---------------|-----------------------------------|
| emr-7.2.0                  | 0.14.1-amzn-1 | Not available.                    |
| emr-5.36,2                 | 0.10.1-amzn-1 | Not available.                    |

| Amazon EMR Release-Etikett | Hudi-Version  | Mit Hudi installierte Komponenten |
|----------------------------|---------------|-----------------------------------|
| 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.                    |
| 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.                    |

| Amazon EMR Release-Etikett | Hudi-Version            | Mit Hudi installierte Komponenten |
|----------------------------|-------------------------|-----------------------------------|
| 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.                    |
| 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.                    |

| Amazon EMR Release-Etikett | Hudi-Version     | Mit Hudi installierte Komponenten |
|----------------------------|------------------|-----------------------------------|
| emr-5.28.1                 | 0.5.0-incubating | Not available.                    |
| emr-5.28.0                 | 0.5.0-incubating | Not available.                    |

# Hue

Hue (Hadoop User Experience) ist eine webbasierte, grafische Open-Source-Benutzeroberfläche zur Verwendung mit Amazon EMR und Apache Hadoop. Hue gruppiert verschiedene Projekte aus dem Hadoop-Ökosystem in einer konfigurierbaren Schnittstelle. Amazon EMR hat auch spezielle Anpassungen für Hue in Amazon EMR hinzugefügt. Hue fungiert als Front-End für Anwendungen, die auf Ihrem Cluster ausgeführt werden, und ermöglicht Ihnen die Interaktion mit Anwendungen über eine bekanntere oder benutzerfreundlichere Benutzeroberfläche. Durch die Anwendungen in Hue, wie beispielsweise Hive- und Pig-Editoren, entfällt das Anmelden im Cluster, um Skripts interaktiv an der Shell der jeweiligen Anwendung auszuführen. Nachdem ein Cluster gestartet wurde, können Sie mit Hue oder einer ähnlichen Schnittstelle vollständig mit Anwendungen interagieren. Weitere Informationen zu Hue finden Sie unter <http://gethue.com>.

Hue wird standardmäßig installiert, wenn Sie Ihren Cluster über die EMR Amazon-Konsole starten. Sie können wählen, ob Sie Hue nicht installieren möchten, indem Sie die erweiterten Optionen in der EMR Amazon-Konsole verwenden, wenn Sie einen Cluster starten, oder indem Sie die `--applications` Option explizit angeben und Hue weglassen, wenn Sie es `create-cluster` von der verwenden. AWS CLI

## Themen

- [Die Hue Versionsinformationen](#)
- [Unterstützte und nicht unterstützte Funktionen von Hue bei Amazon EMR](#)
- [Überlegungen](#)
- [Verbindung zur Hue-Webbenutzeroberfläche herstellen](#)
- [Hue mit einer Remote-Datenbank in Amazon verwenden RDS](#)
- [Erweiterte Konfigurationen für Hue](#)
- [Hue-Versionsverlauf](#)

## Die Hue Versionsinformationen

### Hue-Version für 7.2.0

In der folgenden Tabelle sind die Version von Hue aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Hue EMR installiert.

Die Version der Komponenten, die in dieser Version mit Hue installiert wurden, finden Sie unter [Komponentenversionen der Version 7.2.0](#).

#### Hue-Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                       |
|---------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

#### Hue-Version für 6.15.0

In der folgenden Tabelle sind die Version von Hue aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Hue EMR installiert.

Die Version der Komponenten, die mit Hue in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Hue-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                     |
|---------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| EMRAmazon-Freigabeetikett | Hue-Version | Mit Hue installierte Komponenten                                                                                         |
|---------------------------|-------------|--------------------------------------------------------------------------------------------------------------------------|
|                           |             | nager, hadoop-yarn-resourcenmanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

## Hue-Version für 5.36.2

In der folgenden Tabelle sind die Version von Hue aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Hue EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Hue installiert wurden, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

## Hue-Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|---------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcenmanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |



# Unterstützte und nicht unterstützte Funktionen von Hue bei Amazon EMR

Mit Amazon EMR 7.0.0 und höher benötigt Hue Python 3.9 oder höher. Wenn Sie die Verschlüsselung während der Übertragung verwenden, benötigen Sie einen 2048-Bit-Verschlüsselungsschlüssel.

- Amazon S3 und Hadoop File System (HDFS) Browser
  - Mit den entsprechenden Berechtigungen können Sie Daten zwischen dem kurzlebigen HDFS Speicher und den S3-Buckets, die zu Ihrem Konto gehören, durchsuchen und zwischen diesen verschieben.
  - Standardmäßig können Superuser in Hue auf alle Dateien zugreifen, auf die EMR IAM Amazon-Rollen zugreifen dürfen. Neu erstellte Benutzer verfügen nicht automatisch über Zugriffsberechtigungen auf den Amazon-S3-Filebrowser und für ihre Gruppen müssen die `filebrowser.s3_access`-Berechtigungen aktiviert sein.
- Hive – Führen Sie interaktive Abfragen für Ihre Daten aus. Hive kann außerdem für die programmgesteuerte Prototyp-Erstellung und für als Stapel verarbeitete Abfragen genutzt werden.
- Pig – Führen Sie Skripts für Ihre Daten aus oder nutzen Sie interaktive Befehle.
- Oozie – Erstellen und überwachen Sie Oozie-Workflows.
- Metastore-Manager – Mit dem Metastore-Manager können Sie die Inhalte des Hive-Metastores anzeigen und bearbeiten (importieren/erstellen, löschen usw.).
- Auftragsbrowser – Sehen Sie sich den Status Ihrer gesendeten Hadoop-Aufträge an.
- Benutzerverwaltung — Verwalte Hue-Konten und integriere LDAP Benutzer in Hue.
- AWS Beispiele — Es gibt mehrere "ready-to-run" Beispiele für die Verarbeitung von Beispieldaten aus verschiedenen AWS Diensten mithilfe von Anwendungen in Hue. Wenn Sie sich bei Hue anmelden, gelangen Sie zur Startseite der Hue-Anwendung, wo sich die vorinstallierten Beispiele befinden.
- Livy Server wird nur in EMR Amazon-Version 5.9.0 und höher unterstützt.
- Um Hue-Notebook für Spark zu verwenden, müssen Sie Hue mit Livy und Spark installieren.
- Trino/Presto – Hue unterstützt die Verbindung zu Trino- und Presto-Datenservern. Mit Amazon EMR 7.0.0 und höher wird diese Konnektivität durch `presto-python-client` UND-Konnektoren erleichtert. `trino-python-client` Versionen unter 7.0.0 verwenden einen JDBC Connector über ein JDBC Gateway.

- Das Hue-Dashboard wird nicht unterstützt.
- Postgre SQL wird nicht unterstützt.

## Überlegungen

Beachten Sie die folgenden Einschränkungen, wenn Sie [Hue](#) bei Amazon verwendenEMR.

### Leistung bei großen Hue-Metadatentabellen

Wenn die Hue-Metadaten-Datenbank zu groß wird, kann sich die Leistung verschlechtern. Um die Größe der Tabellen zu überprüfen, stellen Sie zunächst eine Verbindung zum primären Knoten des Amazon EMR on EC2 Amazon-Clusters her SSH und führen Sie den Befehl `sudo mysql -u root`, um My zu starten SQLCLI. Führen Sie die Abfrage aus, um die Größe Ihrer Tabelle zu ermitteln `SELECT COUNT(*) FROM hue.<table_name>`. Wofür sehen Sie im Folgenden `<table_name>` kann sein:

- desktop\_dokument
- Desktop\_Dokument2
- oozie\_job
- beeswax\_savedquery
- bienenwachs\_session
- Geschichte von beeswax\_query

Wenn die Ausführung dieser Abfrage eine Anzahl von mehr als 100000 zurückgibt, sollten Sie den folgenden Bereinigungsbefehl ausführen, um die alten Datensätze zu löschen.

```
cd /opt/cloudera/parcels/CDH/lib/hue # Hue home directory ./build/env/bin/hue
desktop_document_cleanup
```

Weitere Informationen zur Bereinigung Ihrer Datenbank finden Sie auf der [Referenzseite in der Hue-Dokumentation](#).

Hue bereinigt die Tabellen nicht automatisch, aber EMR Amazon-Versionen 5.12.0 und höher bieten eine Methode zum Löschen alter Dokumente in den Tabellen. Erstellen Sie das folgende Shell-Skript und führen Sie es als Schritt in einem EMR Amazon-Cluster mit einem Integer-Parameter aus, der

die maximale Anzahl von Tagen angibt, für die Dokumente in der Metadaten-Datenbank aufbewahrt werden sollen.

```
#!/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
```

## Inkompatibilität zwischen Hue-Versionen

Wenn du dieselbe Hue-Metadaten-Datenbank für mehrere Hue-fähige Cluster verwendest, empfehlen wir, dass auf diesen Clustern dieselbe Version von Hue ausgeführt wird. Verschiedene Versionen von Hue können unterschiedliche Schemas für die Hue-Metadaten-Datenbank haben. Die Verwendung derselben Datenbank für verschiedene Versionen kann dazu führen, dass eine Hue-Installation fehlschlägt. Wenn Sie beispielsweise versuchen, dieselbe Datenbank für zwei Cluster zu verwenden, auf denen 4.10.0 und 4.11.0 installiert sind, kann dies zu Anmeldefehlern für Benutzer führen, die versuchen, sich beim 4.10.0-Cluster anzumelden.

Eine Liste der EMR Amazon-Release-Labels und der entsprechenden installierten Versionen von [Hue finden Sie in der Hue-Versionshistorie](#).

## Verbindung zur Hue-Webbenutzeroberfläche herstellen

Die Verbindung zur Hue-Weboberfläche entspricht der Verbindung zu einer beliebigen HTTP Schnittstelle, die auf dem Master-Knoten eines Clusters gehostet wird. Im folgenden Verfahren wird beschrieben, wie Sie auf die Hue-Benutzeroberfläche zugreifen. Weitere Informationen finden Sie unter [Auf EMR Clustern gehostete Weboberflächen](#) anzeigen im Amazon EMR Management Guide.

So zeigen Sie die Hue-Webbenutzeroberfläche an

1. Folgen Sie diesen Anweisungen zur [Einrichtung eines SSH Tunnels zum Master-Knoten mithilfe der dynamischen Portweiterleitung](#) im Amazon EMR Management Guide.
2. Geben Sie die folgende Adresse in Ihren Browser ein, um die Hue-Weboberfläche zu öffnen: `http://master public DNS:8888` wo *master public dns* ist zum Beispiel der öffentliche DNS Name Ihres Cluster-Masterknotensec2-11-22-333-44.compute-1.amazonaws.com.
3. Wenn Sie sich zum ersten Mal als Administrator anmelden, geben Sie im Hue-Anmeldebildschirm einen Benutzernamen und ein Passwort ein, um Ihr Hue-Super-Konto einzurichten, und wählen Sie dann Konto erstellen aus. Geben Sie andernfalls Ihren Benutzernamen und Ihr Passwort ein, und wählen Sie Create account (Konto erstellen) aus bzw. geben Sie die Anmeldeinformationen ein, die Ihnen von Ihrem Administrator bereitgestellt wurden.

## Hue mit einer Remote-Datenbank in Amazon verwenden RDS

Standardmäßig werden Hue-Benutzerinformationen und Abfrageverläufe in einer lokalen SQL My-Datenbank auf dem Master-Knoten gespeichert. Alternativ können Sie einen oder mehrere HUE-fähige Cluster mithilfe einer in Amazon S3 gespeicherten Konfiguration und einer Meine SQL Datenbank in Amazon Relational Database Service (Amazon) erstellen. RDS Auf diese Weise können Sie Benutzerinformationen und den von Hue erstellten Abfrageverlauf beibehalten, ohne Ihren EMR Amazon-Cluster am Laufen zu halten. Wir empfehlen die Verwendung der serverseitigen Amazon-S3-Verschlüsselung zum Speichern der Konfigurationsdatei.

Erstellen Sie zunächst die Remote-Datenbank für Hue.


Um die externe Datenbank „MeineSQL“ zu erstellen

1. Öffnen Sie die RDS Amazon-Konsole unter <https://console.aws.amazon.com/rds/>.
2. Klicken Sie auf Launch a DB Instance (Eine DB-Instance starten).
3. Wählen Sie Mein SQL und klicken Sie auf Auswählen.
4. Behalten Sie die Standardauswahl für Multi-AZ Deployment und Provisioned IOPS Storage bei und klicken Sie auf Weiter.
5. Übernehmen Sie die Standardwerte für die Instance-Spezifikationen, geben Sie Einstellungen an, und klicken Sie auf Next (Weiter).

6. Wählen Sie auf der Seite "Configure Advanced Settings" eine geeignete Sicherheitsgruppe und einen Datenbanknamen aus. Die von Ihnen verwendete Sicherheitsgruppe muss mindestens den TCP Eingangszugriff für Port 3306 vom Master-Knoten Ihres Clusters aus zulassen. Wenn Sie zu diesem Zeitpunkt noch keinen Cluster erstellt haben, können Sie für alle Hosts eine Verbindung mit Port 3306 zulassen und die Sicherheitsgruppe anpassen, nachdem Sie den Cluster gestartet haben. Klicken Sie auf Launch DB Instance (DB-Instance starten).
7. Wählen Sie im RDS Dashboard Instances und dann die Instanz aus, die Sie gerade erstellt haben. Wenn Ihre Datenbank verfügbar ist, notieren Sie sich den Datenbanknamen, den Benutzernamen, das Passwort und den RDS Instanz-Hostnamen. Diese Informationen benötigen Sie zum Erstellen und Konfigurieren Ihres Clusters.


Um eine externe My SQL database für Hue anzugeben, wenn Sie einen Cluster starten, verwenden Sie AWS CLI

Um beim Starten eines Clusters mithilfe von eine externe My SQL database für Hue anzugeben AWS CLI, verwenden Sie die Informationen, die Sie bei der Erstellung Ihrer RDS Instanz für die Konfiguration `hue.ini` mit einem Konfigurationsobjekt notiert haben

 Note

Sie können mehrere Cluster mit derselben externen Datenbank erstellen, aber jeder Cluster gibt Abfrageverläufe und Benutzerdaten frei.

- Erstellen Sie mit dem einen Cluster AWS CLI, auf dem Hue installiert ist, verwenden Sie die externe Datenbank, die Sie erstellt haben, und verweisen Sie auf eine Konfigurationsdatei mit einer Konfigurationsklassifizierung für Hue, die die Datenbankeigenschaften spezifiziert. Im folgenden Beispiel wird ein Cluster mit installiertem Hue erstellt, der auf die Konfigurationsdatei Amazon S3 in `myConfig.json` verweist, die die Datenbankkonfiguration angibt.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Beispielinhalte der `myConfig.json`-Datei werden unten angezeigt. Ersetzen *dbname*, *username*, *password*, und *RDS instance hostname* mit den Werten, die Sie zuvor im RDS Dashboard notiert haben.

```
[{
 "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": []
 }
]
 }
]
}]
```

## Fehlerbehebung

Im Falle eines RDS Amazon-Failovers

Beim Ausführen einer Abfrage können Verzögerungen auftreten, da die Hue-Datenbank-Instance nicht reagiert oder sich in einem Failover befindet. Im Folgenden finden Sie einige Fakten und Leitlinien für dieses Problem:

- Wenn Sie sich bei der RDS Amazon-Konsole anmelden, können Sie nach Failover-Ereignissen suchen. Um beispielsweise festzustellen, ob ein Failover ausgeführt wird oder aufgetreten ist, suchen Sie nach Ereignissen wie "Multi-AZ instance failover started" und "Multi-AZ instance failover completed".
- Es dauert etwa 30 Sekunden, bis eine RDS Instance einen Failover abgeschlossen hat.
- Wenn Sie länger-als-normal Antworten auf Anfragen in Hue erhalten, versuchen Sie, die Abfrage erneut auszuführen.

## Erweiterte Konfigurationen für Hue

In diesem Abschnitt werden folgende Themen behandelt.

Themen

- [Konfigurieren Sie Hue für Benutzer LDAP](#)

### Konfigurieren Sie Hue für Benutzer LDAP

Die Integration mit LDAP ermöglicht es Benutzern, sich mit vorhandenen Anmeldeinformationen, die in einem LDAP Verzeichnis gespeichert sind, bei Hue anzumelden. Wenn Sie Hue mit integrierenLDAP, müssen Sie Benutzerinformationen in Hue nicht eigenständig verwalten. Die folgenden Informationen zeigen die Hue-Integration mit Microsoft Active Directory, aber die Konfigurationsoptionen sind analog zu jedem LDAP Verzeichnis.

LDAPDie Authentifizierung verbindet sich zunächst mit dem Server und stellt die Verbindung her. Anschließend wird die hergestellte Verbindung für alle nachfolgenden Abfragen verwendet, um nach LDAP Benutzerinformationen zu suchen. Wenn Ihr Active Directory-Server keine anonymen Verbindungen zulässt, muss eine Verbindung mit einem definierten Bind-Namen und -Passwort hergestellt werden. Der definierte Bind-Name (DN) wird durch die Konfigurationseinstellung `bind_dn` definiert. Das Bind-Passwort wird von der Konfigurationseinstellung `bind_password` definiert. Hue bietet zwei Möglichkeiten, LDAP Anfragen zu binden: Suchbindung und Direktbindung. Die bevorzugte Methode für die Verwendung von Hue mit Amazon EMR ist Search Bind.

Wenn Search Bind mit Active Directory verwendet wird, zieht Hue das Benutzernamenattribut (definiert durch `user_name_attr config`) heran, um das Attribut zu finden, das vom definierten Base-Namen (DN) abgerufen werden muss. Search Bind ist hilfreich, wenn der vollständige DN für den Hue-Benutzer nicht bekannt ist.

Es ist beispielsweise möglich, dass Sie `user_name_attr config` auf den allgemeinen Namen (oder CN) festgelegt haben. In diesem Fall verwendet der Active Directory-Server den Hue-Benutzernamen, der während der Anmeldung angegeben wurde, um die Verzeichnisstruktur nach einem übereinstimmenden allgemeinen Namen zu durchsuchen, und zwar beginnend mit dem definierten Base-Namen. Wenn er den allgemeinen Namen für den Hue-Benutzer findet, gibt er den definierten Namen zurück. Hue erstellt dann einen definierten Namen, der für die Authentifizierung des Benutzers durch eine Bind-Operation verwendet wird.

#### Note

Search Bind durchsucht die Benutzernamen in allen Verzeichnis-Unterstrukturen beginnend mit dem definierten Namen. Der in der LDAP Hue-Konfiguration angegebene definierte Basisname sollte dem Benutzernamen am nächsten stehen, da sonst Ihre LDAP Authentifizierungsleistung beeinträchtigt werden kann.

Wenn Direct Bind mit Active Directory verwendet wird, muss die exakte `nt_domain` oder `ldap_username_pattern` für die Authentifizierung verwendet werden. Wenn bei der Verwendung von Direct Bind das NT-Domainattribut (definiert von der Konfigurationseinstellung `nt_domain`) definiert wird, wird eine Vorlage für den definierten Benutzernamen im folgenden Format erstellt: `<login username>@nt_domain`. Diese Vorlage wird verwendet, um alle Verzeichnis-Unterstrukturen beginnend mit dem definierten Namen zu durchsuchen. Wenn die NT-Domain nicht konfiguriert ist, sucht Hue nach einem exakten DN-Muster für den Benutzer (definiert durch die Konfigurationseinstellung `ldap_username_pattern`). In diesem Fall sucht der Server nach einem passenden `ldap_username_pattern`-Wert in allen Verzeichnis-Unterstrukturen beginnend mit dem definierten Base-Namen.

Um einen Cluster mit LDAP Eigenschaften für Hue zu starten, verwenden Sie AWS CLI

- Um LDAP Eigenschaften für anzugeben `hue -ini`, erstellen Sie einen Cluster, auf dem Hue installiert ist, und verweisen Sie auf eine JSON-Datei mit Konfigurationseigenschaften für LDAP. Unten wird ein Beispielbefehl angezeigt, der auf eine Konfigurationsdatei `myConfig.json` verweist, die Amazon S3 gespeichert ist.



```
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.
```

Beispielinhalte von `myConfig.json` werden unten angezeigt.

```
[
 {
 "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

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Sie tun dies, indem Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Um die LDAP Einstellungen in Hue anzusehen

1. Stellen Sie sicher, dass Sie über eine aktive VPN Verbindung oder einen aktiven SSH Tunnel zum Master-Knoten des EMR Amazon-Clusters verfügen. Geben Sie dann in Ihrem Browser Folgendes ein *master-public-dns*:8888, um die Hue-Weboberfläche zu öffnen.
2. Melden Sie sich mit Ihren Hue-Administratoranmeldeinformationen an. Klicken Sie im Dialogfeld Did you know? (Wussten Sie schon?) auf Got it, prof! (Verstanden!) um es zu schließen.
3. Klicken Sie auf das Symbol Hue in der Symbolleiste.
4. Wählen Sie auf der Seite About Hue (Über Hue) auf Configuration (Konfiguration) aus.
5. Klicken Sie im Abschnitt Configuration Sections and Variables (Konfigurationsabschnitte und -variablen) auf Desktop.
6. Scrollen Sie zum Abschnitt LDAP, um Ihre Einstellungen anzuzeigen.

## Hue-Versionsverlauf

Die folgende Tabelle listet die Version von Hue auf, die in jeder Release-Version von Amazon EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Hue Versionsinformationen

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                       |
|----------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.2.0                  | 4.11.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-5.36,2                 | 4.10.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-7.0.0                  | 4.11.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.11.0                 | 4.11.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.1                 | 4.10.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.9.0                  | 4.10.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.1                  | 4.10.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.8.0                  | 4.10.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.7.0                  | 4.10.0      | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.3.1                  | 4.9.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.3.0                  | 4.9.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.2.1                  | 4.8.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.2.0                  | 4.8.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.1.1                  | 4.7.1       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.0                  | 4.7.1       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-6.0.1                  | 4.4.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.0.0                  | 4.4.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-5.33.0                 | 4.9.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.1                 | 4.8.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| emr-5.31.0                 | 4.7.1       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.2                 | 4.6.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                        |
|----------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server   |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| emr-5.28.0                 | 4.4.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.27.1                 | 4.4.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| emr-5.14.0                 | 4.1.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.13.1                 | 4.1.0       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| emr-5.9.1                  | 4.0.1       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.9.0                  | 4.0.1       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                      |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |
| 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-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                                                     |
|----------------------------|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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                              |
| 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                              |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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               |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                          |
|----------------------------|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                                        |
|----------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                                          |
|----------------------------|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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                    |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                       |
|----------------------------|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Amazon EMR Release-Etikett | Hue-Version | Mit Hue installierte Komponenten                                                                                                                                                                                       |
|----------------------------|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| emr-4.1.0                  | 3.7.1       | emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server |



# Iceberg

[Apache Iceberg](#) ist ein offenes Tabellenformat für große Datensätze in Amazon Simple Storage Service (Amazon S3). Es bietet schnelle Abfrageleistung bei großen Tabellen, atomare Commits, gleichzeitige Schreibvorgänge und eine kompatible Tabellenentwicklung. SQL Ab Amazon EMR 6.5.0 können Sie Apache Spark 3 auf EMR Amazon-Clustern mit dem Iceberg-Tabellenformat verwenden.

In der folgenden Tabelle sind die Version von Iceberg aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Iceberg EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Iceberg installiert wurden, finden Sie unter Komponentenversionen von Version 7.2.0.](#)

Iceberg-Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Iceberg-Version      | Mit Iceberg installierte Komponenten |
|---------------------------|----------------------|--------------------------------------|
| emr-7.2.0                 | Iceberg 1.5.0-amzn-0 | Not available.                       |

In der folgenden Tabelle sind die Version von Iceberg aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Iceberg EMR installiert.

Die Version der Komponenten, die mit Iceberg in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Iceberg-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Iceberg-Version      | Mit Iceberg installierte Komponenten |
|---------------------------|----------------------|--------------------------------------|
| emr-6.15.0                | Iceberg 1.4.0-amzn-0 | Not available.                       |

Themen

- [Funktionsweise von Iceberg](#)

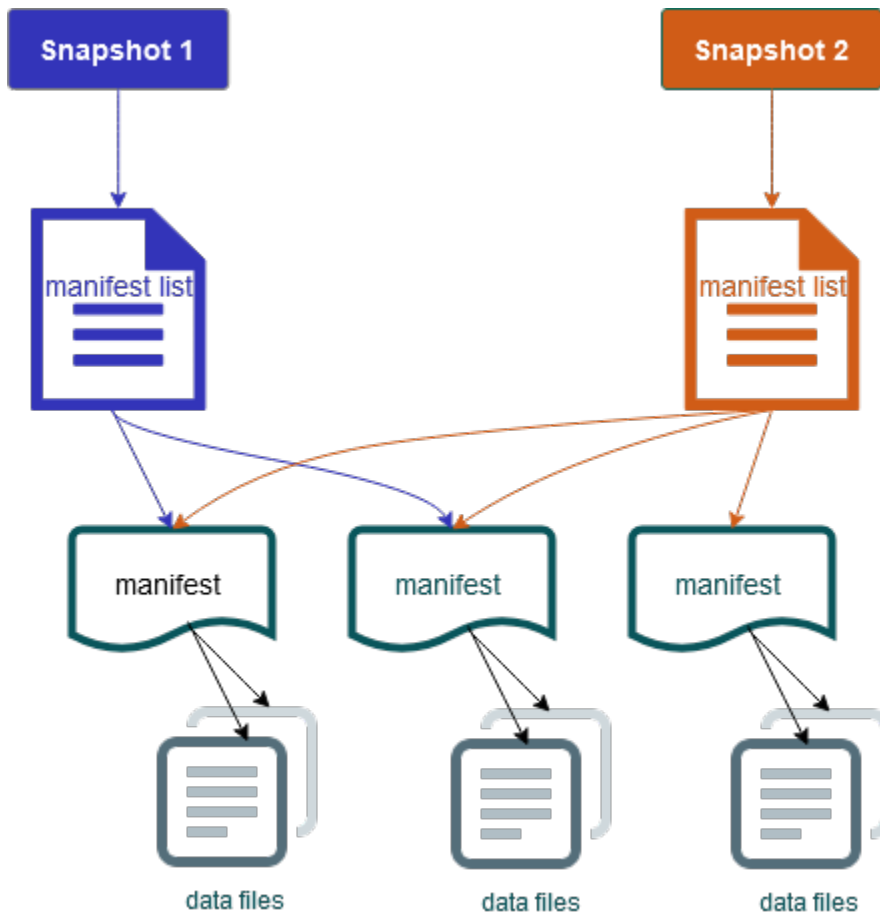
- [Einen Cluster verwenden, auf dem Iceberg installiert ist](#)
- [Überlegungen und Einschränkungen bei der Verwendung von Iceberg bei Amazon EMR](#)
- [Versionsverlauf von Iceberg](#)

## Funktionsweise von Iceberg

Iceberg verfolgt einzelne Datendateien in einer Tabelle statt in Verzeichnissen. Auf diese Weise können Autoren Datendateien an Ort und Stelle erstellen (Dateien werden nicht verschoben oder geändert). Außerdem können Autoren der Tabelle nur in einem expliziten Commit Dateien hinzufügen. Der Tabellenstatus wird in Metadatendateien beibehalten. Bei allen Änderungen am Tabellenstatus wird eine neue Metadatendatei erstellt, die die älteren Metadaten atomar ersetzt. Die Tabellenmetadatendatei verfolgt das Tabellenschema, die Partitionierungskonfiguration und andere Eigenschaften.

Sie enthält auch Snapshots des Tabelleninhalts. Jeder Snapshot ist ein vollständiger Satz von Datendateien in der Tabelle zu einem bestimmten Zeitpunkt. Snapshots sind in der Metadatendatei aufgeführt, aber die Dateien in einem Snapshot werden in separaten Manifestdateien gespeichert. Die atomaren Übergänge von einer Tabellenmetadatendatei zur nächsten ermöglichen die Isolierung von Snapshots. Leser verwenden den Snapshot, der aktuell war, als sie die Tabellenmetadaten geladen haben. Leser sind von Änderungen erst betroffen, wenn sie den Vorgang aktualisieren und einen neuen Speicherort für Metadaten auswählen. Datendateien in Snapshots werden in einer oder mehreren Manifestdateien gespeichert, die eine Zeile für jede Datendatei in der Tabelle, ihre Partitionsdaten und ihre Metriken enthalten. Ein Snapshot ist die Vereinigung aller Dateien in ihren Manifesten. Manifestdateien können auch von Snapshots gemeinsam genutzt werden, um zu vermeiden, dass Metadaten, die sich selten ändern, neu geschrieben werden.

Iceberg-Snapshot-Diagramm



Iceberg bietet folgende Features:

- Unterstützt ACID Transaktionen und Zeitreisen in Ihrem Amazon S3 S3-Datensee.
- Commit-Wiederholungen profitieren von den Leistungsvorteilen [optimistischer Parallelität](#).
- Die Konfliktlösung auf Dateiebene führt zu einer hohen Parallelität.
- Mit Min-Max-Statistiken pro Spalte in Metadaten können Sie Dateien überspringen, was die Leistung bei selektiven Abfragen steigert.
- Sie können Tabellen in flexiblen Partitionslayouts organisieren, wobei die Partitionsentwicklung Aktualisierungen von Partitionsschemas ermöglicht. Abfragen und Datenvolumen können sich dann ändern, ohne auf physische Verzeichnisse angewiesen zu sein.
- Unterstützt die [Weiterentwicklung und Durchsetzung von Schemas](#).
- Iceberg-Tabellen dienen als idempotente Senken und widerspielbare Quellen. Dies ermöglicht Streaming und Batch-Support mit exakt einmal verwendeten Pipelines. Idempotente Senken verfolgen Schreiboperationen, die in der Vergangenheit erfolgreich waren. Daher kann die Senke

im Falle eines Fehlers erneut Daten anfordern und Daten löschen, wenn sie mehrfach gesendet wurden.

- Zeigen Sie Verlauf und Herkunft an, einschließlich Tabellenentwicklung, Operationsverlauf und Statistiken für jeden Commit.
- Migrieren Sie von einem vorhandenen Datensatz mit einer Auswahl an Datenformaten (ParquetORC, Avro) und Analyse-Engine (Spark, Trino, PrestoDB, Flink, Hive).

## Einen Cluster verwenden, auf dem Iceberg installiert ist

Dieser Abschnitt enthält Informationen zur Verwendung von Iceberg mit Spark, Trino, Flink und Hive.

### Verwenden Sie einen Iceberg-Cluster mit Spark

Ab EMR Amazon-Version 6.5.0 können Sie Iceberg mit Ihrem Spark-Cluster verwenden, ohne dass Bootstrap-Aktionen erforderlich sind. Für EMR Amazon-Versionen 6.4.0 und früher können Sie eine Bootstrap-Aktion verwenden, um alle erforderlichen Abhängigkeiten vorab zu installieren.

In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem Amazon EMR Spark-Cluster zu arbeiten. Um mit der Konsole einen Cluster mit installiertem Iceberg zu erstellen, folgen Sie den Schritten unter [Einen Apache Iceberg Data Lake mit Amazon AthenaEMR, Amazon und Glue erstellen](#). AWS

### Erstellen Sie einen Iceberg-Cluster

Sie können mit dem, dem AWS CLI oder Amazon EMR API einen Cluster mit installiertem Iceberg erstellen. AWS Management Console In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem EMR Amazon-Cluster zu arbeiten. Um mit der Konsole einen Cluster mit installiertem Iceberg zu erstellen, folgen Sie den Schritten unter [Einen Apache Iceberg Data Lake mit Amazon AthenaEMR, Amazon und Glue erstellen](#). AWS

Um Iceberg on Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe von finden Sie AWS CLI unter [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie](#) oder. [Geben Sie eine Konfiguration mit Java anSDK, wenn Sie einen Cluster erstellen](#)

1. Erstellen Sie eine Datei `configurations.json`, mit folgendem Inhalt:

```
[{
```

```
"Classification":"iceberg-defaults",
 "Properties":{"iceberg.enabled":"true"}
}]
```

2. Im nächsten Schritt erstellen Sie einen Cluster mit der folgenden Konfiguration. Ersetzen Sie den Amazon-S3-Beispiel-Bucket-Pfad und die Subnetz-ID durch Ihre eigene.

```
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
```

Alternativ können Sie einen EMR Amazon-Cluster mit der Spark-Anwendung erstellen und die Datei `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` als JAR Abhängigkeit in einen Spark-Job aufnehmen. Weitere Informationen finden Sie unter [Anwendungen einreichen](#).

Um das JAR als Abhängigkeit in einen Spark-Auftrag aufzunehmen, fügen Sie der Spark-Anwendung die folgende Konfigurationseigenschaft hinzu:

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

Weitere Informationen zu den Abhängigkeiten von Spark-Aufträgen finden Sie unter [Abhängigkeitsverwaltung](#) im Apache-Spark-Dokument [Ausführen von Spark in Kubernetes](#).

## Initialisieren Sie eine Spark-Sitzung für Iceberg

Die folgenden Beispiele zeigen, wie Sie die interaktive Spark-Shell starten, Spark Submit verwenden oder Amazon EMR Notebooks verwenden, um mit Iceberg auf Amazon EMR zu arbeiten.

spark-shell

1. Connect zum Master-Knoten her mit SSH. Weitere Informationen finden Sie unter [Connect dem Master-Knoten herstellen SSH](#) im Amazon EMR Management Guide.

2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. Um die PySpark Shell zu verwenden, `spark-shell` ersetzen Sie durch `spark`.

```
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. Connect zum Master-Knoten her mit SSH. Weitere Informationen finden Sie unter [Connect dem Master-Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie den folgenden Befehl ein, um die Spark-Sitzung für Iceberg zu starten.

```
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

Um eine Spark-Sitzung mit EMR Studio-Notebooks zu initialisieren, konfigurieren Sie Ihre Spark-Sitzung mit dem `%%configure` magischen Befehl in Ihrem EMR Amazon-Notizbuch, wie im folgenden Beispiel. Weitere Informationen finden Sie unter [Use EMR Notebooks Magic](#) im 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/"
}
}
```

## Schreiben Sie an einen Iceberg-Tabelle

Das folgende Beispiel zeigt, wie Sie einen Iceberg-Datensatz erstellen DataFrame und ihn als Iceberg-Datensatz schreiben. Die Beispiele zeigen, wie Sie mit Datensätzen arbeiten, indem Sie die Spark-Shell verwenden, während Sie mit dem Master-Knoten SSH als Standard-Hadoop-Benutzer verbunden sind.

### Note

Um Codebeispiele in die Spark-Shell einzufügen, geben Sie an der Eingabeaufforderung `:paste` ein, fügen das Beispiel ein und drücken dann CTRL+D.

## PySpark

Spark umfasst auch eine auf Python basierende Shell namens `pyspark`, mit der Sie Prototypen von in Python geschriebenen Spark-Programmen entwickeln können. `Invoke` wird in `pyspark` auf dem Hauptknoten festgelegt.

```
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()
```

## Lesen von einer Iceberg-Tabelle

### 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
```



## Spark-Eigenschaften so konfigurieren, dass der AWS Glue-Datenkatalog als Metastore für Iceberg-Tabellen verwendet wird

Um den AWS Glue-Katalog als Metastore für Iceberg-Tabellen zu verwenden, stellen Sie die Spark-Konfigurationseigenschaften wie folgt ein:

```
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
```

## Verwenden Sie einen Iceberg-Cluster mit Trino

Ab EMR Amazon-Version 6.6.0 können Sie Iceberg mit Ihrem Trino-Cluster verwenden.

In diesem Tutorial verwenden Sie den AWS CLI um mit Iceberg auf einem Amazon EMR Trino-Cluster zu arbeiten. Um mit der Konsole einen Cluster mit installiertem Iceberg zu erstellen, folgen Sie den Schritten unter [Einen Apache Iceberg Data Lake mit Amazon AthenaEMR, Amazon und Glue erstellen](#). AWS

### Erstellen Sie einen Iceberg-Cluster

Um Iceberg on Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe von finden Sie AWS CLI unter [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie](#) oder [Geben Sie eine Konfiguration mit Java an SDK, wenn Sie einen Cluster erstellen](#)

1. Erstellen Sie eine `iceberg.properties`-Datei und legen Sie einen Wert für den ausgewählten Katalog fest. Wenn Sie beispielsweise den Hive-Metastore als Ihren Katalog verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben.

```
connector.name=iceberg
hive.metastore.uri=thrift://localhost:9083
```

Wenn Sie den AWS Glue-Datenkatalog als Ihren Shop verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben.

```
connector.name=iceberg
iceberg.catalog.type=glue
```

- Erstellen Sie eine Bootstrap-Aktion, die `iceberg.properties` von Amazon S3 nach `/etc/trino/conf/catalog/iceberg.properties` kopiert, wie im folgenden Beispiel. Weitere Informationen zu Bootstrap-Aktionen finden Sie unter [Erstellen von Bootstrap-Aktionen zum Installieren zusätzlicher Software](#).

```
set -ex
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/iceberg.properties
```

- Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie dabei den Beispiel-Skriptpfad und den Schlüsselnamen für Bootstrap-Aktionen durch Ihren eigenen.

```
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>
```

## Initialisieren Sie eine Trino-Sitzung für Iceberg

Um eine Trino-Sitzung zu initialisieren, führen Sie den folgenden Befehl aus.

```
trino-cli --catalog iceberg
```

## Schreiben Sie an einen Iceberg-Tabelle

Erstellen Sie mit den folgenden SQL Befehlen Ihre Tabelle und schreiben Sie in sie.

```
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');
```

## Lesen von Iceberg aus einer Tabelle

Um von Ihrer Iceberg-Tabelle zu lesen, führen Sie den folgenden Befehl aus.

```
trino> SELECT * from default.iceberg_table;
```

## Verwenden Sie einen Iceberg-Cluster mit Flink

Ab EMR Amazon-Version 6.9.0 können Sie Iceberg mit einem Flink-Cluster verwenden, ohne dass die Einrichtungsschritte erforderlich sind, wenn Sie die Open-Source-Iceberg Flink-Integration verwenden.

## Erstellen Sie einen Iceberg-Cluster

Sie können einen Cluster mit installiertem Iceberg mithilfe von Amazon AWS Management Console, AWS CLI, The oder Amazon EMR API erstellen. In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem EMR Amazon-Cluster zu arbeiten. Um mit der Konsole einen Cluster mit installiertem Iceberg zu erstellen, folgen Sie den Schritten unter [Einen Apache Iceberg Data Lake mit Amazon AthenaEMR, Amazon und Glue](#) erstellen. AWS

Um Iceberg on Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe von finden Sie AWS CLI unter [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie](#) oder. [Geben Sie eine Konfiguration mit Java an SDK, wenn Sie einen Cluster erstellen](#) Erstellen Sie eine Datei mit dem Namen `configurations.json` und dem folgenden Inhalt:

```
[{
```

```
"Classification":"iceberg-defaults",
 "Properties":{"iceberg.enabled":"true"}
}]
```

Erstellen Sie als Nächstes einen Cluster mit der folgenden Konfiguration und ersetzen Sie den Amazon-S3-Bucket-Beispielpfad und die Subnetz-ID durch Ihre eigenen Werte:

```
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
```

Alternativ können Sie einen Amazon EMR 6.9.0-Cluster mit einer darin enthaltenen Flink-Anwendung erstellen und die Datei `/usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar` als JAR Abhängigkeit in einem Flink-Job verwenden.

## Verwenden des Flink-Clients SQL

Das SQL Client-Skript befindet sich unter `/usr/lib/flink/bin`. Sie können das Skript mit dem folgenden Befehl ausführen:

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode
./sql-client.sh
```

Dadurch wird eine SQL Flink-Shell gestartet.

## Flink-Beispiele

Erstellen Sie eine Iceberg-Tabelle

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);

```

## Tabelle 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);");

```

## Schreiben Sie an einen Iceberg-Tabelle

### Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

## Tabelle API

```
tEnv.executeSql(
 "INSERT INTO `glue_catalog`.`"
 + db
 + "`.`sample` values (1, 'a'),(2,'b'),(3,'c');");
```

## Datenstrom 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");
```

## Lesen von einer Iceberg-Tabelle

### Blinzeln SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

### Tabelle API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`");
```

### Datenstrom 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");
```

## Verwenden des Hive-Katalogs

Stellen Sie sicher, dass die Abhängigkeiten zwischen Flink und Hive wie unter [Konfigurieren Sie Flink mit Hive Metastore und Glue Catalog](#) beschrieben aufgelöst wurden.

## Flink-Auftrag ausführen

Eine Möglichkeit, einen Job an Flink zu senden, besteht darin, eine Flink-Sitzung pro Job zu verwenden. YARN Dies kann mit dem folgenden Befehl gestartet werden:

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

## Verwenden Sie einen Iceberg-Cluster mit Hive

Mit EMR Amazon-Versionen 6.9.0 und höher können Sie Iceberg mit einem Hive-Cluster verwenden, ohne die Einrichtungsschritte durchführen zu müssen, die für die Open Source Iceberg Hive-Integration erforderlich sind. Für EMR Amazon-Versionen 6.8.0 und früher können Sie eine Bootstrap-Aktion verwenden, um `iceberg-hive-runtime` JAR für die Konfiguration von Hive für die Iceberg-Unterstützung zu installieren.

Amazon EMR 6.9.0 beinhaltet alle Funktionen für die [Integration von Hive 3.1.3 mit Iceberg 0.14.1](#) und beinhaltet auch EMR zusätzliche Funktionen von Amazon wie die auto Auswahl unterstützter Ausführungs-Engines zur Laufzeit (Amazon auf 6.9.0). EMR EKS

## Erstellen Sie einen Iceberg-Cluster

Sie können mit dem, dem AWS CLI oder Amazon EMR API einen Cluster mit installiertem Iceberg erstellen. AWS Management Console In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem EMR Amazon-Cluster zu arbeiten. Um mit der Konsole einen Cluster mit installiertem Iceberg zu erstellen, folgen Sie den Schritten unter [Build an Iceberg Data Lake using Amazon AthenaEMR, Amazon und Glue](#). AWS

Um Iceberg on Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifikation mithilfe von AWS CLI oder Java SDK finden Sie unter [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie](#) oder. [Geben Sie eine Konfiguration mit Java an SDK, wenn Sie](#)



[einen Cluster erstellen](#) Erstellen Sie eine Datei mit dem Namen `configurations.json` und dem folgenden Inhalt:

```
[{
 "Classification":"iceberg-defaults",
 "Properties":{"iceberg.enabled":"true"}
}]
```

Erstellen Sie als Nächstes einen Cluster mit der folgenden Konfiguration und ersetzen Sie dabei den Amazon-S3-Bucket-Beispielpfad und die Subnetz-ID durch Ihre eigene:

```
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
```

Ein Hive-Iceberg-Cluster macht die folgenden Dinge:

- Lädt das Iceberg Hive-Laufzeit-Jar in Hive und aktiviert die Iceberg-bezogene Konfiguration für die Hive-Engine.
- Aktiviert die dynamische Ausführungs-Engine-Auswahl von Amazon EMR Hive, um zu verhindern, dass Benutzer die unterstützte Ausführungs-Engine für die Iceberg-Kompatibilität einrichten.

#### Note

Hive Iceberg-Cluster unterstützen AWS Glue Data Catalog derzeit nicht. Der Standard-Iceberg-Katalog ist `HiveCatalog`, was dem für die Hive-Umgebung konfigurierten Metastore entspricht. Weitere Informationen zur Katalogverwaltung finden Sie unter [Verwenden HCatalog](#) in der [Apache Hive-Dokumentation](#).

## Feature-Unterstützung

Amazon EMR 6.9.0 unterstützt Hive 3.1.3 und Iceberg 0.14.1. Die Featureunterstützung ist auf Iceberg-kompatible Feature für Hive 3.1.2 und 3.1.3 beschränkt. Die folgenden -Befehle sind verfügbar:

- Bei den EMR Amazon-Versionen 6.9.0 bis 6.12.x müssen Sie das `libfb303` JAR in das Hive-Verzeichnis aufnehmen. `auxlib` Verwenden Sie den folgenden Befehl, um es einzubinden:

```
sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/libfb303.jar
```

Bei EMR Amazon-Versionen 6.13 und höher wird das `libfb303` JAR automatisch mit dem `auxlib` Hive-Verzeichnis verknüpft.

- Erstellen einer Tabelle
  - Nicht partitionierte Tabelle – Externe Tabellen in Hive können erstellt werden, indem der Storage-Handler wie folgt bereitgestellt wird:

```
CREATE EXTERNAL TABLE x (i int) STORED BY
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

- Partitionierte Tabelle – Externe partitionierte Tabellen in Hive können wie folgt erstellt werden:

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

### Note

Das `STORED AS DateiformatORC/AVRO/PARQUET` wird in Hive 3 nicht unterstützt. Die Standardoption und einzige Option ist Parquet.

- Tabelle löschen – Der `DROP TABLE`-Befehl wird verwendet, um Tabellen zu löschen, wie im folgenden Beispiel:

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- Eine Tabelle lesen – `SELECT`-Anweisungen können verwendet werden, um Iceberg-Tabellen in Hive zu lesen, wie im folgenden Beispiel. Unterstützte Ausführungs-Engines sind MR und Tez.

```
SELECT * FROM table_name
```

[Informationen zur Select-Syntax von Hive finden Sie unter LanguageManual Select.](#) Informationen zu Select-Anweisungen mit Iceberg-Tabellen in Hive finden Sie unter [Apache Iceberg Select](#).

- In eine Tabelle einfügen – Die INSERT INTO-HiveQL-Anweisung funktioniert nur für Iceberg-Tabellen mit Unterstützung für die Map-Reduce-Ausführungseingine. EMRAmazon-Benutzer müssen die Ausführungseingine nicht explizit festlegen, da Amazon EMR Hive die Eingine für Iceberg Tables zur Laufzeit auswählt.
- Einzelne Tabelle einfügen in – Beispiel:

```
INSERT INTO table_name VALUES ('a', 1);
INSERT INTO table_name SELECT...;
```

- Einfügen in mehrere Tabellen – Das Einfügen in Anweisungen, die nicht aus mehreren Tabellen bestehen, werden unterstützt. Beispiel:

```
FROM source
INSERT INTO table_1 SELECT a, b
INSERT INTO table_2 SELECT c,d;
```

## Überlegungen und Einschränkungen bei der Verwendung von Iceberg bei Amazon EMR

Dieser Abschnitt enthält Überlegungen und Einschränkungen für die Verwendung von Iceberg mit Spark, Trino, Flink und Hive.

### Überlegungen zur Verwendung von Iceberg mit Spark

- Amazon EMR 6.5.0 unterstützt Iceberg, das auf Amazon EMR on läuft, EKS standardmäßig nicht. Ein benutzerdefiniertes Amazon EMR 6.5.0-Image ist verfügbar, sodass Sie es `--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` als `spark-submit` Parameter zur Erstellung von Iceberg-Tabellen auf Amazon EMR übergeben können. EKS Weitere Informationen finden Sie unter [Einreichen eines Spark-Workloads in Amazon EMR mithilfe eines benutzerdefinierten Images](#) im Amazon EMR on EKS Development Guide. Sie können sich

auch an AWS Support wenden, um Unterstützung zu erhalten. Ab Amazon EMR 6.6.0 wird Iceberg auf Amazon EMR unterstützt. EKS

- Wenn Sie AWS Glue als Katalog für Iceberg verwenden, stellen Sie sicher, dass die Datenbank, in der Sie eine Tabelle erstellen, in AWS Glue vorhanden ist. Wenn Sie Dienste wie verwenden AWS Lake Formation und den Katalog nicht laden können, stellen Sie sicher, dass Sie über den richtigen Zugriff auf den Dienst verfügen, um den Befehl auszuführen.

## Überlegungen zur Verwendung von Iceberg mit Trino

- Amazon EMR 6.5 bietet keine native Trino Iceberg Catalog-Unterstützung für Iceberg. Trino benötigt Iceberg v0.11, daher empfehlen wir, einen EMR Amazon-Cluster für Trino getrennt vom Spark-Cluster zu starten und Iceberg v0.11 in diesem Cluster zu integrieren.
- Wenn Sie AWS Glue als Katalog für Iceberg verwenden, stellen Sie sicher, dass die Datenbank, in der Sie eine Tabelle erstellen, in AWS Glue vorhanden ist. Wenn Sie Dienste wie verwenden AWS Lake Formation und den Katalog nicht laden können, stellen Sie sicher, dass Sie über den richtigen Zugriff auf den Dienst verfügen, um den Befehl auszuführen.

## Überlegungen zur Verwendung von Iceberg mit Flink

Wenn Sie AWS Glue als Katalog für Iceberg verwenden, stellen Sie sicher, dass die Datenbank, in der Sie eine Tabelle erstellen, in AWS Glue vorhanden ist. Wenn Sie Dienste wie verwenden AWS Lake Formation und den Katalog nicht laden können, stellen Sie sicher, dass Sie über den richtigen Zugriff auf den Dienst verfügen, um den Befehl auszuführen.

## Überlegungen zur Verwendung von Iceberg mit Hive

- Iceberg unterstützt die folgenden Abfragetypen:
  - Create table
  - Drop table
  - In die Tabelle einfügen
  - Tabelle lesen
- Nur die MR (MapReduce) -Ausführungseingine wird für Operationen DML (Datenmanipulationssprache) unterstützt, und MR ist in Hive 3.1.3 veraltet.

- AWS Glue Data Catalog wird derzeit nicht für Iceberg with Hive unterstützt.
- Die Fehlerbehandlung ist nicht robust genug. Im Falle einer Fehlkonfiguration können Einfügungen in Abfragen erfolgreich abgeschlossen werden. Wenn Metadaten nicht aktualisiert werden, kann das jedoch zu Datenverlust führen.

## Versionsverlauf von Iceberg

In der folgenden Tabelle sind die Version von Iceberg aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

### Informationen zur Iceberg-Version

| Amazon EMR Release-Etikett | Iceberg-Version | Mit Iceberg installierte Komponenten |
|----------------------------|-----------------|--------------------------------------|
| emr-7.2.0                  | 1.5.0-amzn-0    | Not available.                       |
| emr-7.1.0                  | 1.4.3-amzn-0    | Not available.                       |
| emr-7.0.0                  | 1.4.2-amzn-0    | Not available.                       |
| emr-6.15.0                 | 1.4.0-amzn-0    | Not available.                       |
| emr-6.14.0                 | 1.3.1-amzn-0    | Not available.                       |
| emr-6.13.0                 | 1.3.0-amzn-1    | Not available.                       |
| emr-6.12.0                 | 1.3.0-amzn-0    | Not available.                       |
| emr-6.11.1                 | 1.2.0-amzn-0    | Not available.                       |
| emr-6.11.0                 | 1.2.0-amzn-0    | Not available.                       |
| emr-6.10.1                 | 1.1.0-amzn-0    | Not available.                       |
| emr-6.10.0                 | 1.1.0-amzn-0    | Not available.                       |

| Amazon EMR Release-Etikett | Iceberg-Version | Mit Iceberg installierte Komponenten |
|----------------------------|-----------------|--------------------------------------|
| emr-6.9.1                  | 0.14.1-amzn-0   | Not available.                       |
| emr-6.9.0                  | 0.14.1-amzn-0   | Not available.                       |
| emr-6.8.1                  | 0.14.0-amzn-0   | Not available.                       |
| emr-6.8.0                  | 0.14.0-amzn-0   | Not available.                       |
| emr-6.7.0                  | 0.13.1-amzn-0   | Not available.                       |
| emr-6.6.0                  | 0.13.1          | Not available.                       |
| emr-6.5.0                  | 0.12.0          | Not available.                       |

## Iceberg-Versionshinweise nach Version

- [Amazon EMR 6.9.0 — Versionshinweise zu Iceberg](#)

### Amazon EMR 6.9.0 — Versionshinweise zu Iceberg

#### Amazon EMR 6.9.0 — Änderungen bei Iceberg

| Typ      | Beschreibung                                                                                                                       |
|----------|------------------------------------------------------------------------------------------------------------------------------------|
| Funktion | Amazon EMR Flink-Integration mit Iceberg.                                                                                          |
| Funktion | Amazon EMR Hive-Integration mit Iceberg.                                                                                           |
| Funktion | Support für das Zwischenspeichern von Iceberg-Metadatendateien auf Amazon FSx for Lustre, um die Abfrageplanungszeit zu verkürzen. |
| Backport | <a href="#">PR 5050</a> : Flink 1.15: Support Schreiboptionen in den Inline-Einfügekomentaren. SQL                                 |

| Typ      | Beschreibung                                                                                                                                                                                                      |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Backport | <a href="#">PR 5282</a> : AWS: Beheben Sie PUT Wiederholungsfehler, indem Sie neue Datendateistreams öffnen.                                                                                                      |
| Backport | <a href="#">PR 5318</a> : Flink 1.15: Überbrückt die Lücke zwischen FlinkSource und IcebergSource (FLIP-27) und es wurde eine Opt-in-Konfiguration hinzugefügt, um die -27-Quelle in Flink zu verwenden. FLIP SQL |
| Backport | <a href="#">PR 5344</a> : Flink 1.14: Überbrückt die Lücke zwischen FlinkSource und IcebergSource (FLIP-27) und es wurde eine Opt-in-Konfiguration hinzugefügt, um die -27-Quelle in Flink zu verwenden. FLIP SQL |
| Backport | <a href="#">PR 5393</a> : Flink 1.14, 1.15: Vermeiden Sie die Konvertierung von Iceberg in Flink-Metriken im -27 Source Reader. MetricContext FLIP                                                                |
| Backport | <a href="#">PR 5401</a> : Flink 1.14, 1.15: Fehlende Gruppe in PR #5393 für -27 Quellleser-Metriken. IcebergSourceReader FLIP                                                                                     |
| Backport | <a href="#">PR 5679</a> : Spark 3.2, 3.3: Die Übertragung der NULL-Zulässigkeit für den Knoten wurde korrigiert. MergeRows                                                                                        |
| Backport | <a href="#">PR 5860</a> : Spark 3.3: Fix QueryFailure bei der Ausführung RewriteManifestProcedure auf datumpartitionierten Tabellen.                                                                              |
| Backport | <a href="#">PR 5880</a> : Spark 3.3: NULL-Zulässigkeit in Projektionen behoben. merge-on-read                                                                                                                     |
| Backport | <a href="#">PR 5917</a> : Spark 3.2: NULL-Zulässigkeit in Projektionen korrigieren. merge-on-read                                                                                                                 |

# Jupyter Notebook in Amazon EMR

[Jupyter Notebook](#) ist eine Open-Source-Webanwendung, mit der Sie Dokumente erstellen und freigeben können, die Live-Code, Gleichungen, Visualisierungen und narrativen Text enthalten. Amazon EMR bietet Ihnen drei Möglichkeiten, mit Jupyter Notebooks zu arbeiten:

Themen

- [EMR Studio](#)
- [Amazon EMR Notebook auf Basis von Jupyter Notebook](#)
- [JupyterHub](#)

## EMR Studio

Amazon EMR Studio ist eine webbasierte integrierte Entwicklungsumgebung (IDE) für vollständig verwaltete [Jupyter Notebooks](#), die auf Amazon-EMR-Clustern ausgeführt werden. Sie können ein EMR Studio für Ihr Team einrichten, um in R, Python, Scala und PySpark geschriebene Anwendungen zu entwickeln, zu visualisieren und zu debuggen.

Wir empfehlen die Verwendung von EMR Studio, wenn Sie Jupyter Notebooks auf Amazon EMR verwenden. Weitere Informationen finden Sie unter [EMR Studio](#) im Verwaltungshandbuch für Amazon EMR.

## Amazon EMR Notebook auf Basis von Jupyter Notebook

EMR Notebooks ist eine in die Amazon-EMR-Konsole integrierte [Jupyter-Notebook-Umgebung](#), mit der Sie schnell Jupyter Notebooks erstellen, sie an Spark-Cluster anhängen und dann den Jupyter-Notebook-Editor in der Konsole öffnen können, um Abfragen und Code remote auszuführen. Ein EMR Notebook wird unabhängig von Clustern in Amazon S3 gespeichert, um dauerhaften Speicher, schnellen Zugriff und Flexibilität zu gewährleisten. Sie können mehrere Notebooks öffnen, mehrere Notebooks an einen einzelnen Cluster anhängen und ein Notizbuch in verschiedenen Clustern wiederverwenden.

Weitere Informationen finden Sie unter [Nutzung von EMR Notebooks](#) im Verwaltungshandbuch für Amazon EMR.

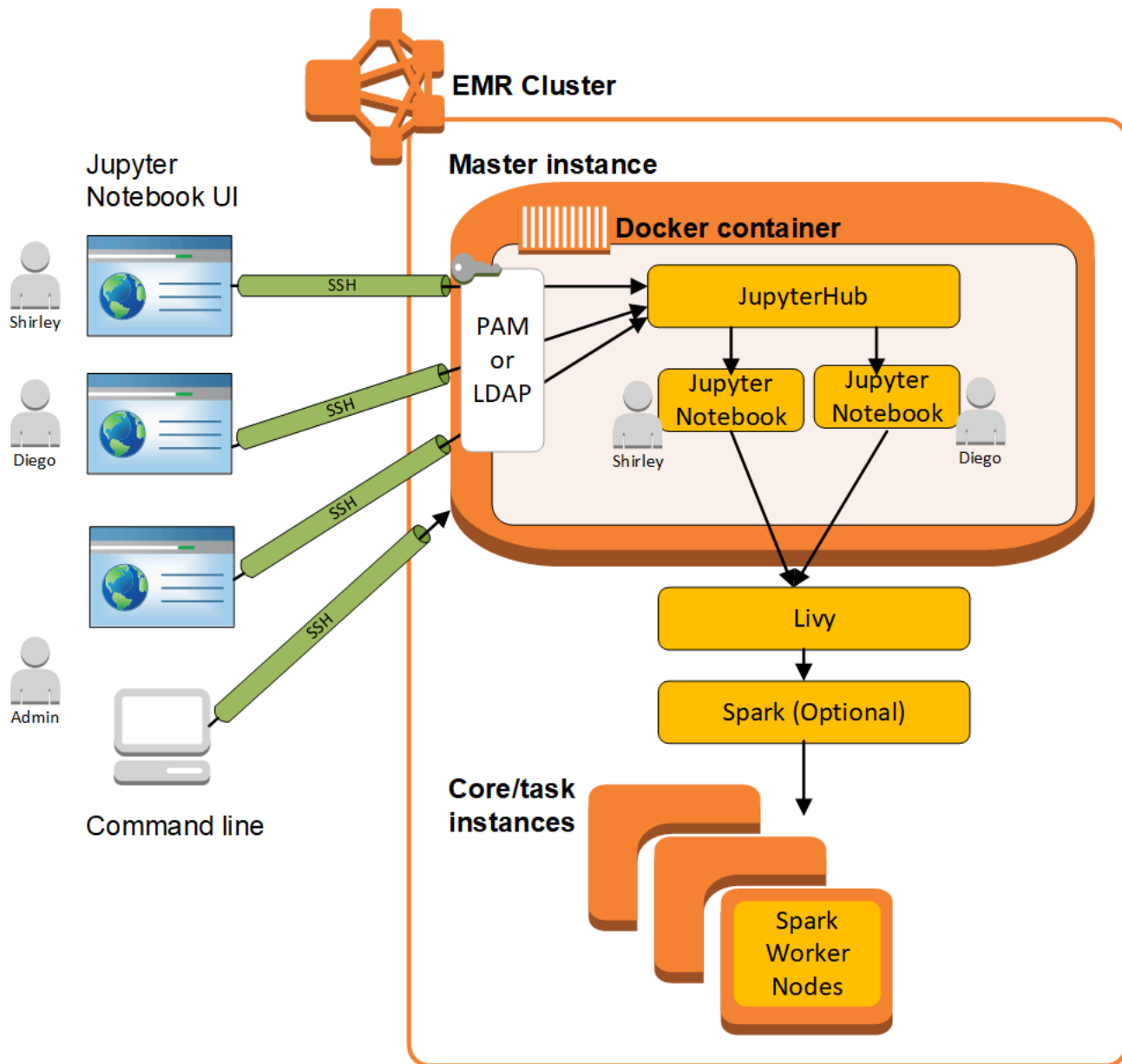


# JupyterHub

[Jupyter Notebook](#) ist eine Open-Source-Webanwendung, mit der Sie Dokumente erstellen und freigeben können, die Live-Code, Gleichungen, Visualisierungen und narrativen Text enthalten. [JupyterHub](#) ermöglicht es Ihnen, mehrere Instanzen eines Jupyter-Notebookservers für einen einzelnen Benutzer zu hosten. Wenn Sie einen Cluster mit erstellen JupyterHub, EMR erstellt Amazon einen Docker-Container auf dem Master-Knoten des Clusters. JupyterHub, alle für Jupyter und [Sparkmagic](#) erforderlichen Komponenten werden innerhalb des Containers ausgeführt.

Sparkmagic ist eine Kernelbibliothek, die es Jupyter-Notebooks ermöglicht, mit [Apache Spark](#) zu interagieren, der auf Amazon EMR läuft [Apache Livy](#), einem Server für Spark. REST Spark und Apache Livy werden automatisch installiert, wenn Sie einen Cluster mit erstellen. JupyterHub Der Standard-Python-3-Kernel für Jupiter ist zusammen mit den Kernen PySpark 3,, und Spark verfügbar PySpark, die mit Sparkmagic verfügbar sind. Sie können diese Kernel verwenden, um Ad-hoc-Spark-Code und interaktive SQL Abfragen mit Python und Scala auszuführen. Sie können zusätzliche Kernel innerhalb des Docker-Containers manuell installieren. Weitere Informationen finden Sie unter [Installieren von zusätzlicher Kernen und Bibliotheken](#).

Das folgende Diagramm zeigt die Komponenten von JupyterHub on Amazon EMR mit den entsprechenden Authentifizierungsmethoden für Notebook-Benutzer und den Administrator. Weitere Informationen finden Sie unter [Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren](#).



In der folgenden Tabelle sind die Version von JupyterHub aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. JupyterHub

Informationen zur Version der Komponenten, mit denen JupyterHub in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

## JupyterHub Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | JupyterHub Ausführung | Komponenten, die mit installiert wurden 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 |

In der folgenden Tabelle sind die Version von JupyterHub aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. JupyterHub

Informationen zur Version der Komponenten, mit denen JupyterHub in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 6.15.0](#).

## JupyterHub Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | JupyterHub Ausführung | Komponenten, die mit installiert wurden 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, |

| EMRAmazon-Freigabeetikett | JupyterHub Ausführung | Komponenten, die mit installiert wurden JupyterHub                                                                                             |
|---------------------------|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
|                           |                       | hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub |

In der folgenden Tabelle sind die Version von JupyterHub aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. JupyterHub

Informationen zur Version der Komponenten, mit denen JupyterHub in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

JupyterHub Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | JupyterHub Ausführung | Komponenten, die mit installiert wurden 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 |

Der Python-3-Kernel, JupyterHub der bei Amazon enthalten ist, EMR ist 3.6.4.

Die im `jupyterhub` Container installierten Bibliotheken können zwischen EMR Amazon-Release-Versionen und EC2 AMI Amazon-Versionen variieren.

Installierte Bibliotheken mit **conda** auflisten

- Führen Sie den folgenden Befehl in der Befehlszeile des Master-Knotens aus:

```
sudo docker exec jupyterhub bash -c "conda list"
```

Installierte Bibliotheken mit **pip** auflisten

- Führen Sie den folgenden Befehl in der Befehlszeile des Master-Knotens aus:

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

Themen

- [Erstellen Sie einen Cluster mit JupyterHub](#)
- [Überlegungen bei der Verwendung JupyterHub bei Amazon EMR](#)
- [Konfiguration JupyterHub](#)
- [Konfigurieren von Persistenz für Notebooks in Amazon S3](#)
- [Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen](#)
- [JupyterHub Konfiguration und Verwaltung](#)
- [Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren](#)
- [Installieren von zusätzlicher Kernen und Bibliotheken](#)
- [JupyterHub Versionsverlauf](#)

## Erstellen Sie einen Cluster mit JupyterHub

Sie können einen EMR Amazon-Cluster JupyterHub mithilfe von AWS Management Console AWS Command Line Interface, oder Amazon erstellen EMR API. Stellen Sie sicher, dass der Cluster nicht mit der Option zum automatischen Beenden nach Abschluss der Schritte angelegt wird (Option `--auto-terminate` in der AWS CLI). Stellen Sie außerdem sicher, dass Administratoren und Notebook-Benutzer auf das Schlüsselpaar zugreifen können, das Sie beim Erstellen des

Clusters verwenden. Weitere Informationen finden Sie unter [Verwenden eines key pair für SSH Anmeldeinformationen](#) im Amazon EMR Management Guide.

## Erstellen Sie JupyterHub mithilfe der Konsole einen Cluster

Gehen Sie wie folgt vor, um einen Cluster mit JupyterHub installierten erweiterten Optionen in der EMR Amazon-Konsole zu erstellen.

Um einen EMR Amazon-Cluster zu erstellen, der über die EMR Amazon-Konsole JupyterHub installiert ist

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Unter Software Configuration (Softwarekonfiguration):
  - Wählen Sie für Release emr-5.36.2 aus und wählen Sie. JupyterHub
  - Wenn Sie Spark verwenden, um den AWS Glue-Datenkatalog als Metastore für Spark zu verwenden SQL, wählen Sie Für Spark-Tabellenmetadaten verwenden aus. Weitere Informationen finden Sie unter [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).
  - Wählen Sie unter Softwareeinstellungen bearbeiten die Option Konfiguration eingeben und Werte angeben, oder wählen Sie JSON aus S3 laden und geben Sie eine JSON Konfigurationsdatei an. Weitere Informationen finden Sie unter [Konfiguration JupyterHub](#).
4. Konfigurieren Sie unter Add steps (optional) (Schritte hinzufügen (optional)) die Schritte, die ausgeführt werden sollen, wenn der Cluster erstellt wird, stellen Sie sicher, dass Auto-terminate cluster after the last step is completed (Cluster automatisch beenden, nachdem der letzte Schritt ausgeführt wurde) nicht ausgewählt ist, und klicken Sie auf Next (Weiter).
5. Wählen Sie die Option Hardware Configuration (Hardwarekonfiguration), Next (Weiter). Weitere Informationen finden [Sie unter Konfiguration von Cluster-Hardware und Netzwerken](#) im Amazon EMR Management Guide.
6. Wählen Sie Optionen für General Cluster Settings (Allgemeine Cluster-Einstellungen), Next (Weiter).
7. Wählen Sie Security Options (Sicherheitsoptionen), geben Sie ein Schlüsselpaar an und wählen Sie Create Cluster (Cluster erstellen).

## Erstellen Sie einen Cluster JupyterHub mit AWS CLI

Um einen Cluster mit zu starten JupyterHub, verwenden Sie den `aws emr create-cluster` Befehl und geben Sie für die `--applications` Option Folgendes an `Name=JupyterHub`. Das folgende Beispiel startet einen JupyterHub Cluster auf Amazon EMR mit zwei EC2 Instances (eine Master- und eine Core-Instance). Außerdem ist das Debugging aktiviert, wobei die Protokolle am Amazon-S3-Speicherort gespeichert werden wie in `--log-uri` angegeben. Das angegebene `key pair` ermöglicht den Zugriff auf EC2 Amazon-Instances im Cluster.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

## Überlegungen bei der Verwendung JupyterHub bei Amazon EMR

Beachten Sie bei der Verwendung bei JupyterHub Amazon Folgendes EMR.

### Warning

Benutzer-Notebooks und Dateien werden in dem Dateisystem auf dem Master-Knoten gespeichert. Hierbei handelt es sich um flüchtigen Speicher, der nach der Beendigung des Clusters nicht dauerhaft bestehen bleibt. Wenn ein Cluster beendet wird, gehen diese Daten verloren, wenn sie nicht gesichert sind. Wir empfehlen, dass Sie regelmäßige Sicherungen mithilfe von `cron`-Jobs oder anderen für Ihre Anwendung geeigneten Methoden einplanen.

Außerdem bleiben Konfigurationsänderungen innerhalb des Containers bei einem Neustart des Containers möglicherweise nicht bestehen. Wir empfehlen Ihnen, die Containerkonfiguration per Skript vorzunehmen oder anderweitig zu automatisieren, damit Sie Anpassungen leichter reproduzieren können.

- Die Kerberos-Authentifizierung, die mit einer EMR Amazon-Sicherheitskonfiguration eingerichtet wurde, wird nicht unterstützt.
- [OAuthenticator](#) wird nicht unterstützt.

## Konfiguration JupyterHub

Sie können die Konfiguration von Notebooks bei JupyterHub Amazon EMR und einzelnen Benutzern anpassen, indem Sie eine Verbindung zum Cluster-Masterknoten herstellen und die Konfigurationsdateien bearbeiten. Nachdem Sie Werte geändert haben, starten Sie den `jupyterhub`-Container.

Ändern Sie die Eigenschaften in den folgenden Dateien, um einzelne Jupyter-Notebooks zu konfigurieren JupyterHub :

- `jupyterhub_config.py` – Diese Datei befindet sich standardmäßig im `/etc/jupyter/conf/`-Verzeichnis auf dem Hauptknoten. Weitere Informationen finden Sie in der Dokumentation unter [Grundlagen der Konfiguration](#). JupyterHub
- `jupyter_notebook_config.py` – Diese Datei wird im `/etc/jupyter/`-Verzeichnis gespeichert und standardmäßig in den `jupyterhub`-Container kopiert. Weitere Informationen finden Sie unter [Konfigurationsdatei und Befehlszeilen-Optionen](#) in der Jupyter-Notebook-Dokumentation.

Sie können auch die Konfigurationsklassifizierung `jupyter-sparkmagic-conf` verwenden, um Sparkmagic anzupassen. Dadurch werden Werte in der `config.json`-Datei für Sparkmagic aktualisiert. Weitere Informationen zu verfügbaren Einstellungen finden Sie unter [example\\_config.json](#) unter. GitHub Weitere Informationen zur Verwendung von Konfigurationsklassifizierungen mit Anwendungen in Amazon finden Sie EMR unter [Anwendungen konfigurieren](#).

Im folgenden Beispiel wird ein Cluster mit dem gestartet AWS CLI, der auf die Datei `MyJupyterConfig.json` für die Sparkmagic-Konfigurationsklassifizierungseinstellungen verweist.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).



```
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
```

Beispiele für den Inhalt der Datei `MyJupyterConfig.json`:

```
[
 {
 "Classification": "jupyter-sparkmagic-conf",
 "Properties": {
 "kernel_python_credentials" : "{\"username\":\"diego\", \"base64_password\":
\"mypass\", \"url\":\"http://localhost:8998\", \"auth\":\"None\"}"
 }
 }
]
```

#### Note

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Sie tun dies, indem Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Konfigurieren von Persistenz für Notebooks in Amazon S3

Sie können einen JupyterHub Cluster in Amazon EMR so konfigurieren, dass von einem Benutzer gespeicherte Notizbücher in Amazon S3 bestehen bleiben, außerhalb des kurzlebigen Speichers auf Cluster-Instances. EC2

Sie geben die Amazon-S3-Persistenz mithilfe der `jupyter-s3-conf`-Konfigurationsklassifizierung an, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Zusätzlich zur Aktivierung der Amazon-S3-Persistenz mithilfe der Eigenschaft `s3.persistence.enabled` geben Sie einen Bucket in Amazon S3 an, in dem Notebooks mithilfe

der Eigenschaft `s3.persistence.bucket` gespeichert werden. Notebooks für jeden Benutzer werden in einem `jupyter/jupyterhub-user-name`-Ordner im angegebenen Bucket gespeichert. Der Bucket muss bereits in Amazon S3 vorhanden sein, und die Rolle für das EC2 Instance-Profil, das Sie bei der Erstellung des Clusters angeben, muss über Berechtigungen für den Bucket verfügen (standardmäßig lautet die Rolle `EMR_EC2_DefaultRole`). Weitere Informationen finden [Sie unter IAM Rollen für EMR Amazon-Berechtigungen für AWS Dienste konfigurieren](#).

Wenn Sie einen neuen Cluster mit denselben Konfigurationsklassifizierungseigenschaften verwenden, können Benutzer Notebooks mit den Inhalten des Speicherorts öffnen.

Beachten Sie, dass wenn Sie Dateien als Module in ein Notebook importieren und Amazon S3 aktiviert haben, dies dazu führt, dass die Dateien auf Amazon S3 hochgeladen werden. Wenn Sie Dateien importieren, ohne die Amazon S3 S3-Persistenz zu aktivieren, werden sie in Ihren JupyterHub Container hochgeladen.

Im folgenden Beispiel wird Amazon-S3-Persistenz aktiviert. Von Benutzern gespeicherte Notebooks werden im Ordner `s3://MyJupyterBackups/jupyter/jupyterhub-user-name` der einzelnen Benutzer gespeichert. Dabei ist `jupyterhub-user-name` ein Benutzername wie beispielsweise `diego`.

```
[
 {
 "Classification": "jupyter-s3-conf",
 "Properties": {
 "s3.persistence.enabled": "true",
 "s3.persistence.bucket": "MyJupyterBackups"
 }
 }
]
```

## Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen

JupyterHub Administratoren und Notebook-Benutzer müssen über einen SSH Tunnel eine Verbindung zum Cluster-Masterknoten herstellen und dann eine Verbindung zu den Webschnittstellen herstellen, die JupyterHub auf dem Master-Knoten bereitgestellt werden. Weitere Informationen zur Konfiguration eines SSH Tunnels und zur Verwendung des Tunnels für Proxy-Webverbindungen finden [Sie unter Verbindung zum Cluster](#) herstellen im Amazon EMR Management Guide.

Standardmäßig EMR ist JupyterHub On Amazon über Port 9443 auf dem Master-Knoten verfügbar. Der interne JupyterHub Proxy bedient auch Notebook-Instances über Port 9443. JupyterHub und auf Jupyter-Webschnittstellen kann URL mit einem nach dem folgenden Muster zugegriffen werden:

```
https://MasterNodeDNShttps://----sep----:9443
```

Sie können unter Verwendung der `c.JupyterHub.port`-Eigenschaft in der `jupyterhub_config.py`-Datei einen anderen Port angeben. Weitere Informationen finden Sie in der Dokumentation unter [Netzwerkgrundlagen](#). JupyterHub

Standardmäßig EMR verwendet JupyterHub Amazon ein selbstsigniertes Zertifikat für die SSL Verschlüsselung HTTPS. Benutzer werden aufgefordert, das selbstsignierte Zertifikat als vertrauenswürdig einzustufen, wenn sie eine Verbindung herstellen. Sie können ein vertrauenswürdiges Zertifikat und eigene Schlüssel verwenden. Ersetzen Sie die Standard-Zertifikatsdatei, `server.crt`, und die Schlüsseldatei, `server.key`, im `/etc/jupyter/conf/`-Verzeichnis auf dem Master-Knoten durch Ihre eigenen Zertifikat- und Schlüsseldateien. Verwenden Sie die `c.JupyterHub.ssl_cert` Eigenschaften `c.JupyterHub.ssl_key` und in der `jupyterhub_config.py` Datei, um Ihre SSL Materialien zu spezifizieren. Weitere Informationen finden Sie in der JupyterHub Dokumentation unter [Sicherheitseinstellungen](#). Nachdem Sie `jupyterhub_config.py` geändert haben, starten Sie den Container neu.

## JupyterHub Konfiguration und Verwaltung

JupyterHub und verwandte Komponenten laufen in einem Docker-Container mit dem Namen `jupyterhub`, auf dem das Ubuntu-Betriebssystem ausgeführt wird. Sie haben mehrere Möglichkeiten, Komponenten zu verwalten, die innerhalb des Containers ausgeführt werden.

### Warning

Anpassungen, die Sie innerhalb des Containers vornehmen, bleiben möglicherweise nicht erhalten, wenn der Container neu gestartet wird. Wir empfehlen Ihnen, die Containerkonfiguration per Skript vorzunehmen oder anderweitig zu automatisieren, damit Sie Anpassungen leichter reproduzieren können.

## Administration über die Befehlszeile

Wenn Sie mit dem Master-Knoten verbunden sind SSH, können Sie Befehle ausgeben, indem Sie die Docker-Befehlszeilenschnittstelle (CLI) verwenden und den Container mit Namen (`jupyterhub`)

oder ID angeben. Beispiel: `sudo docker exec jupyterhub command` führt Befehle aus, die vom Betriebssystem oder einer im Container ausgeführten Anwendung erkannt werden. Mit dieser Methode können Sie Benutzer zum Betriebssystem hinzufügen und zusätzliche Anwendungen und Bibliotheken innerhalb des Docker-Containers installieren. Beispiel: Das Standard-Container-Image enthält Conda für die Paketinstallation, sodass Sie den folgenden Befehl in der Befehlszeile des Master-Knotens ausführen können, um eine Anwendung, Keras, innerhalb des Containers zu installieren:

```
sudo docker exec jupyterhub conda install keras
```

## Administration durch Übermitteln von Schritten

Schritte sind eine Möglichkeit, Arbeit an einen Cluster zu übermitteln. Schritte können Sie senden, wenn der Cluster gestartet oder ausgeführt wird. Befehle, die Sie in der Befehlszeile ausführen, können mit `command-runner.jar` als Schritte übermittelt werden. Weitere Informationen finden Sie unter [Arbeiten mit Schritten mithilfe der CLI AND-Konsole](#) im Amazon EMR Management Guide und [Befehle und Skripte auf einem Amazon-EMR-Cluster ausführen](#).

Sie könnten beispielsweise den folgenden AWS CLI Befehl auf einem lokalen Computer verwenden, um Keras auf die gleiche Weise zu installieren, wie Sie es im vorherigen Beispiel über die Befehlszeile des Master-Knotens getan haben:

```
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"
```

Sie können auch eine Folge von Schritten per Skript festlegen, das Skript in Amazon S3 hochladen und dann `script-runner.jar` verwenden, um das Skript auszuführen, wenn Sie den Cluster erstellen, oder das Skript als Schritt hinzufügen. Weitere Informationen finden Sie unter [Befehle und Skripte auf einem Amazon-EMR-Cluster ausführen](#). Ein Beispiel finden Sie unter [the section called "Beispiel: Bash-Skript zum Hinzufügen mehrerer Benutzer"](#).

## Verwaltung mit REST APIs

Jupyter und der HTTP Proxy für JupyterHub Provide REST APIs, mit dem Sie Anfragen senden können. JupyterHub Um Anfragen an zu senden JupyterHub, müssen Sie der Anfrage ein API Token beifügen. Sie können den `curl` Befehl von der Master-Node-Befehlszeile aus verwenden, um REST Befehle auszuführen. Weitere Informationen finden Sie in den folgenden Ressourcen:

- Die [Verwendung JupyterHub findet sich REST API](#) in der Dokumentation für JupyterHub, die Anweisungen zum Generieren von API Tokens enthält
- [Jupyter Notebook-Server auf API GitHub](#)
- [configurable-http-proxy](#) auf GitHub

Das folgende Beispiel zeigt die Verwendung von REST API for JupyterHub, um eine Benutzerliste abzurufen. Der Befehl übergibt ein zuvor generiertes Admin-Token und verwendet den Standardport 9443 für JupyterHub, um die Ausgabe zur besseren Anzeige an [jq weiterzuleiten](#):

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \
-H "Authorization: token $admin_token" | jq .
```

## Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren

Sie können eine von zwei Methoden verwenden, um sich für Benutzer zu authentifizieren, JupyterHub sodass sie Notizbücher erstellen und optional verwalten können. JupyterHub Die einfachste Methode ist die Verwendung JupyterHub des austauschbaren Authentifizierungsmoduls (). PAM Darüber hinaus EMR unterstützt JupyterHub Amazon das [LDAPAuthenticator-Plugin JupyterHub zum](#) Abrufen von Benutzeridentitäten von einem LDAP Server, z. B. einem Microsoft Active Directory-Server. Anweisungen und Beispiele für das Hinzufügen von Benutzern mit den beiden Authentifizierungsmethoden finden Sie in diesem Abschnitt.

JupyterHub EMR hat bei Amazon einen Standardbenutzer mit Administratorrechten. Der Benutzername ist  `jovyan`, das Passwort ist  `jupyter`. Wir empfehlen Ihnen dringend, den Benutzer durch einen anderen Benutzer mit Administratorrechten zu ersetzen. Dies ist mit einem Schritt beim Erstellen des Clusters oder durch das Herstellen einer Verbindung zum Master-Knoten möglich, während der Cluster ausgeführt wird.

### Themen

- [PAMAuthentifizierung wird verwendet](#)
- [LDAPAuthentifizierung wird verwendet](#)
- [Benutzer-Identitätswechsel](#)

## PAMAuthentifizierung wird verwendet

Das Erstellen von PAM Benutzern JupyterHub bei Amazon EMR erfolgt in zwei Schritten. Der erste Schritt besteht darin, dem Betriebssystem, das im Container `jupyterhub` auf dem Master-Knoten ausgeführt wird, Benutzer hinzuzufügen, und für jeden Benutzer ein entsprechendes Benutzer-Stammverzeichnis hinzuzufügen. Der zweite Schritt besteht darin, diese Betriebssystembenutzer als JupyterHub Benutzer hinzuzufügen — ein Vorgang, der als Whitelisting bezeichnet wird. JupyterHub Nachdem ein JupyterHub Benutzer hinzugefügt wurde, kann er sich mit dem verbinden JupyterHub URL und seine Betriebssystemanmeldedaten für den Zugriff angeben.

Wenn sich ein Benutzer anmeldet, JupyterHub wird die Notebook-Server-Instanz für diesen Benutzer geöffnet, die im Home-Verzeichnis des Benutzers auf dem Master-Knoten gespeichert ist `/var/lib/jupyter/home/username`. Wenn eine Notebook-Server-Instanz nicht existiert, JupyterHub wird eine Notebook-Instanz im Home-Verzeichnis des Benutzers erzeugt. In den folgenden Abschnitten wird gezeigt, wie Benutzer einzeln zum Betriebssystem und zu hinzugefügt werden JupyterHub, gefolgt von einem rudimentären Bash-Skript, das mehrere Benutzer hinzufügt.

### Hinzufügen eines Betriebssystembenutzers zum Container

Das folgende Beispiel verwendet den Befehl [useradd](#) innerhalb des Containers, um einen einzelnen Benutzer hinzuzufügen, `diego`, und ein Stammverzeichnis für diesen Benutzer anzulegen. Der zweite Befehl verwendet [chpasswd](#), um `diego` ein Passwort für diesen Benutzer zu erteilen. Befehle werden in der Befehlszeile des Master-Knotens ausgeführt, während die Verbindung über SSH Sie können diese Befehle auch mithilfe eines Schritts ausführen, wie zuvor in [Administration durch Übermitteln von Schritten](#) beschrieben.

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

### Einen JupyterHub Benutzer hinzufügen

Sie können das Admin-Panel in JupyterHub oder im verwenden, REST API um Benutzer und Administratoren oder einfach nur Benutzer hinzuzufügen.

So fügen Sie Benutzer und Administratoren über das Admin-Panel in hinzu JupyterHub

1. Stellen Sie über `https://`eine Connect zum Master-Knoten her SSH und melden Sie sich an `MasterNodeDNS:9443` mit einer Identität, die über Administratorrechte verfügt.
2. Wählen Sie Control Panel (Systemsteuerung), Admin (Administrator).

3. Wählen Sie User (Benutzer), Add Users (Benutzer hinzufügen) oder wählen Sie Admin (Administrator), Add Admins (Administratoren hinzufügen).

Um einen Benutzer hinzuzufügen, verwenden Sie REST API

1. Stellen Sie mithilfe des folgenden Befehls Connect zum Master-Knoten her SSH und verwenden Sie den folgenden Befehl auf dem Master-Knoten, oder führen Sie den Befehl als Schritt aus.
2. Erwerben Sie ein Administrator-Token, um API Anfragen zu stellen, und ersetzen Sie es *AdminToken* im folgenden Schritt mit diesem Token.
3. Verwenden Sie den folgenden Befehl und ersetzen Sie *UserName* mit einem Betriebssystembenutzer, der innerhalb des Containers erstellt wurde.

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/users/UserName"
```

#### Note

Sie werden automatisch als Benutzer JupyterHub ohne Administratorrechte hinzugefügt, wenn Sie sich zum ersten Mal an der JupyterHub Weboberfläche anmelden.

Beispiel: Bash-Skript zum Hinzufügen mehrerer Benutzer

Das folgende Bash-Beispielskript verknüpft die vorherigen Schritte in diesem Abschnitt, um mehrere JupyterHub Benutzer zu erstellen. Das Skript kann direkt auf dem Hauptknoten ausgeführt werden, oder in Amazon S3 hochgeladen und als Schritt ausgeführt werden.

Das Skript erstellt zunächst eine Reihe von Benutzernamen und verwendet den `jupyterhub token` Befehl, um ein API Token für den Standardadministrator `joyvan` zu erstellen. Anschließend erstellt es einen Betriebssystem-Benutzer im `jupyterhub`-Container für jeden Benutzer und weist jedem von ihnen ein anfängliches Passwort zu, das gleich ihrem Benutzernamen ist. Schließlich ruft es die REST API Operation auf, in der die einzelnen Benutzer erstellt werden. JupyterHub Es übergibt das zuvor im Skript generierte Token und leitet die REST Antwort `jq` zur besseren Anzeige weiter.

```
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

```

Speichern Sie das Skript an einem Speicherort in Amazon S3, beispielsweise `s3://mybucket/createjupyterusers.sh`. Anschließend können Sie es mit `script-runner.jar` als Schritt ausführen.

Beispiel: Ausführen des Skripts beim Erstellen eines Clusters (AWS CLI)

#### Note

Linux-Zeilensetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```

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"]

```

Ausführen des Skripts auf einem vorhandenen Cluster (AWS CLI)

#### Note

Linux-Zeilensetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).



```
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"]
```

## LDAPAuthentifizierung wird verwendet

Das Lightweight Directory Access Protocol (LDAP) ist ein Anwendungsprotokoll zum Abfragen und Ändern von Objekten, die Ressourcen wie Benutzern und Computern entsprechen, die in einem LDAP-kompatiblen Verzeichnisdienstanbieter wie Active Directory oder einem Open LDAP Server gespeichert sind. Sie können das [LDAPAuthenticator-Plugin for JupyterHub with](#) bei JupyterHub Amazon LDAP für EMR die Benutzerauthentifizierung verwenden. Das Plugin verarbeitet Anmeldesitzungen für LDAP Benutzer und stellt Benutzerinformationen für Jupyter bereit. Auf diese Weise können Benutzer eine Verbindung zu JupyterHub Notebooks herstellen, indem sie die Anmeldeinformationen für ihre Identitäten verwenden, die auf einem -kompatiblen Server gespeichert sind. LDAP

Die Schritte in diesem Abschnitt führen Sie durch die folgenden Schritte zur Einrichtung und Aktivierung LDAP mithilfe des LDAP Authenticator-Plug-ins für JupyterHub. Sie führen die Schritte aus, während Sie mit der Befehlszeile des Master-Knotens verbunden sind. Weitere Informationen finden Sie unter [Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen](#).

1. Erstellen Sie eine LDAP Konfigurationsdatei mit Informationen zum LDAP Server, z. B. der Host-IP-Adresse, dem Port, den Bindungsnamen usw.
2. Ändern Sie `/etc/jupyter/conf/jupyterhub_config.py`, um das LDAP Authenticator-Plug-In für JupyterHub zu aktivieren.
3. Erstellen Sie ein Skript, das LDAP innerhalb des Containers konfiguriert wird, und führen Sie es aus `jupyterhub`.
4. Fragen Sie LDAP nach Benutzern ab und erstellen Sie dann Basisverzeichnisse innerhalb des Containers für jeden Benutzer. JupyterHub benötigt Home-Verzeichnisse zum Hosten von Notebooks.
5. Führen Sie ein Skript aus, das neu gestartet wird JupyterHub

### Important

Testen Sie vor der Einrichtung Ihre Netzwerkinfrastruktur LDAP, um sicherzustellen, dass der LDAP Server und der Cluster-Masterknoten wie gewünscht kommunizieren können.

TLS verwendet in der Regel Port 389 über eine einfache TCP Verbindung. Wenn Ihre LDAP Verbindung verwendet SSL, SSL ist der bekannte TCP Port für 636.

Erstellen Sie die LDAP Konfigurationsdatei

Das folgende Beispiel verwendet die folgenden Platzhalter-Konfigurationswerte. Ersetzen Sie diese durch Parameter, die Ihrer Implementierung entsprechen.

- Auf dem LDAP Server läuft Version 3 und er ist auf Port 389 verfügbar. Dies ist der SSL Standard-Nicht-Port für LDAP.
- Der grundlegende spezifische Name (DN, Distinguished Name) ist `dc=example, dc=org`.

Erstellen Sie mit einem Texteditor die Datei [ldap.conf](#) mit Inhalt, der etwa wie folgt aussieht.

Verwenden Sie für Ihre LDAP Implementierung geeignete Werte. Ersetzen `host` mit der IP-Adresse oder dem auflösbaren Hostnamen Ihres LDAP Servers.

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

Aktivieren Sie das LDAP Authenticator-Plugin für JupyterHub

Verwenden Sie einen Texteditor, um die Datei `/etc/jupyter/conf/jupyterhub_config.py` zu ändern, und fügen Sie [ldapauthenticator](#)-Eigenschaften hinzu, etwa wie folgt. Ersetzen `host` mit der IP-Adresse oder dem auflösbaren Hostnamen des LDAP Servers. Das Beispiel geht davon aus, dass sich die Benutzerobjekte innerhalb einer Organisationseinheit (ou) mit dem Namen befinden `people`, und verwendet die definierten Namenskomponenten, die Sie zuvor mithilfe von eingerichtet haben `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'
```

## Konfigurieren Sie LDAP innerhalb des Containers

Erstellen Sie mithilfe eines Texteditors ein Bash-Skript mit folgendem Inhalt:

```
#!/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_mkhomedir.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"
```

Speichern Sie das Skript auf dem Master-Knoten und führen Sie es in der Befehlszeile des Master-Knotens aus. Wenn das Skript beispielsweise als `configure_ldap_client.sh` gespeichert ist, machen Sie die Datei ausführbar:

```
chmod +x configure_ldap_client.sh
```

Und führen das Skript aus:

```
./configure_ldap_client.sh
```

## Attribute zur Active Directory hinzufügen

Um jeden Benutzer zu finden und den entsprechenden Eintrag in der Datenbank zu erstellen, benötigt der JupyterHub Docker-Container die folgenden UNIX Eigenschaften für das entsprechende Benutzerobjekt in Active Directory. Weitere Informationen finden Sie im Abschnitt [Wie kann ich die GID UID RFC /2307-Attribute weiter bearbeiten](#), nachdem das Unix-Attributs-Plug-In nicht mehr für das Snap-In „Active Directory-Benutzer und -ComputerMMC“ verfügbar ist? im Artikel [Klarstellung zum Status der Identitätsverwaltung für Unix \(IDMU\) und zur NIS Serverrolle in der technischen Vorschau von Windows Server 2016 und](#) darüber hinaus.

- `homeDirectory`

Dies ist der Speicherort im Stammverzeichnis des Benutzers, das in der Regel `/home/username` ist.

- `gidNumber`

Dies ist ein Wert größer als 60.000, der noch nicht von einem anderen Benutzer verwendet wird. Überprüfen Sie die `etc/passwd`-Datei auf GIDs, die bereits verwendet werden.

- `uidNumber`

Dies ist ein Wert größer als 60000, der noch nicht von einer anderen Gruppe verwendet wird. Überprüfen Sie die `etc/group`-Datei auf UIDs, die bereits verwendet werden.

- `uid`

Das ist dasselbe wie `username`.

## Root-Verzeichnisse für Benutzer erstellen

JupyterHub benötigt Home-Verzeichnisse innerhalb des Containers, um LDAP Benutzer zu authentifizieren und Instanzdaten zu speichern. Das folgende Beispiel zeigt zwei Benutzer, Shirley und Diego, im LDAP Verzeichnis.

Der erste Schritt besteht darin, den LDAP Server mit [ldapsearch](#) nach den Benutzer-ID- und Gruppen-ID-Informationen der einzelnen Benutzer abzufragen, wie im folgenden Beispiel gezeigt, und dabei Folgendes zu ersetzen *host* mit der IP-Adresse oder dem auflösbaren Hostnamen Ihres Servers: LDAP

```
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
```

Der *ldapsearch* Befehl gibt eine LDIF -formatierte Antwort zurück, die für die Benutzer Shirley und Diego der folgenden ähnelt.

```
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
```

Unter Verwendung der Informationen aus der Antwort führen Sie Befehle innerhalb des Containers zum Erstellen eines Stammverzeichnisses für jeden allgemeinen Namen (cn) des Benutzers aus. Verwenden Sie `uidNumber` und `gidNumber`, um das Eigentum für das Stammverzeichnis für diesen Benutzer festzulegen. Die folgenden Beispielbefehle erledigen dies für den Benutzer *shirley*.

```
sudo docker container exec jupyterhub bash -c "mkdir /home/shirley"
sudo docker container exec jupyterhub bash -c "chown -R $uidNumber /home/shirley"
sudo docker container exec jupyterhub bash -c "sudo chgrp -R $gidNumber /home/shirley"
```

### Note

LDAPAuthenticator for JupyterHub unterstützt die lokale Benutzererstellung nicht. Weitere Informationen finden Sie in der [Anmerkung zur Konfiguration LDAP des Authentifikators zur lokalen Benutzererstellung](#).

Um einen lokalen Benutzer manuell zu erstellen, verwenden Sie den folgenden Befehl.

```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/home/shirley:/bin/bash' >> /etc/passwd"
```

Starten Sie den Container JupyterHub neu

Führen Sie den folgenden Befehl aus, um den `jupyterhub`-Container neu zu starten:

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

## Benutzer-Identitätswechsel

Ein Spark-Job, der in einem Jupyter-Notebook ausgeführt wird, durchläuft während seiner Ausführung auf Amazon mehrere Anwendungen. EMR Zum Beispiel werden PySpark drei Codes, die ein Benutzer in Jupyter ausführt, von Sparkmagic empfangen. Sparkmagic verwendet eine HTTP POST Anfrage, um ihn an Livy zu senden, das dann einen Spark-Job erstellt, der auf dem Cluster ausgeführt wird, der ihn verwendet. YARN

Standardmäßig werden YARN Jobs, die auf diese Weise übermittelt wurden, als Benutzer ausgeführt `livy`, unabhängig davon, welcher Benutzer den Job initiiert hat. Durch die Einrichtung

eines Benutzerwechsels können Sie festlegen, dass die Benutzer-ID des Notebook-Benutzers auch der Benutzer ist, der dem YARN Auftrag zugeordnet ist. Von shirley und diego initiierte Aufträge werden nicht dem Benutzer livy zugeordnet. Stattdessen werden die von den Benutzern initiierten Aufträge shirley und diego zugeordnet. Dies erleichtert das Überwachen der Jupyter-Nutzung und das Verwalten von Anwendungen in der Organisation.

Diese Konfiguration wird nur unterstützt, wenn Aufrufe von Sparkmagic an Livy ohne Authentifizierung erfolgen. Anwendungen, die eine Authentifizierung oder einen Proxy-Layer zwischen Hadoop-Anwendungen und Livy (z. B. Apache Knox Gateway) bereitstellen, werden nicht unterstützt. Bei den Schritten zur Konfiguration des Benutzerwechsels in diesem Abschnitt wird davon ausgegangen, dass JupyterHub und Livy auf demselben Master-Knoten ausgeführt werden. Wenn Ihre Anwendung separate Cluster hat, [Schritt 3: Erstellen Sie HDFS Basisverzeichnisse für Benutzer](#) muss diese geändert werden, sodass HDFS Verzeichnisse auf dem Livy-Masterknoten erstellt werden.

Schritte zum Konfigurieren des Benutzer-Identitätswechsels

- [Schritt 1: Livy konfigurieren](#)
- [Schritt 2: Benutzer hinzufügen](#)
- [Schritt 3: Erstellen Sie HDFS Basisverzeichnisse für Benutzer](#)

Schritt 1: Livy konfigurieren

Sie verwenden die Konfigurationsklassifizierungen `livy-conf` und `core-site`, wenn Sie einen Cluster erstellen, um Livy-Benutzer-Identitätswechsel zu ermöglichen (siehe folgendes Beispiel). Speichern Sie die Konfigurationsklassifizierung als JSON und verweisen Sie dann darauf, wenn Sie den Cluster erstellen, oder geben Sie die Konfigurationsklassifizierung inline an. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
[
 {
 "Classification": "livy-conf",
 "Properties": {
 "livy.impersonation.enabled": "true"
 }
 },
 {
 "Classification": "core-site",
 "Properties": {
 "hadoop.proxyuser.livy.groups": "*"
 }
 }
]
```

```

 "hadoop.proxyuser.livy.hosts": "*"
 }
}
]

```

## Schritt 2: Benutzer hinzufügen

Fügen Sie JupyterHub Benutzer mit PAM oder hinzuLDAP. Weitere Informationen finden Sie unter [PAMAuthentifizierung wird verwendet](#) und [LDAPAuthentifizierung wird verwendet](#).

## Schritt 3: Erstellen Sie HDFS Basisverzeichnisse für Benutzer

Sie haben eine Verbindung zum Master-Knoten hergestellt, um Benutzer zu erstellen. Kopieren Sie den unten angegebenen Inhalt, während die Verbindung zum Master-Knoten noch besteht, und speichern Sie ihn in einer Skriptdatei. Das Skript erstellt HDFS Home-Verzeichnisse für jeden JupyterHub Benutzer auf dem Master-Knoten. Das Skript geht davon aus, dass Sie die Standard-Administratorbenutzer-ID verwenden, *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
```

## Installieren von zusätzlicher Kerneln und Bibliotheken

Wenn Sie einen Cluster mit JupyterHub auf Amazon erstellenEMR, werden der Standard-Python-3-Kernel für Jupyter zusammen mit den PySpark und Spark-Kernel für Sparkmagic auf dem Docker-Container installiert. Sie können zusätzliche Kernel installieren. Sie können auch zusätzliche Bibliotheken und Pakete installieren und dann für die entsprechende Shell importieren.

### Installieren eines Kernels

Kernel werden innerhalb des Docker-Containers installiert. Am einfachsten ist dies zu bewerkstelligen, indem Sie ein Bash-Skript mit Installationsbefehlen erstellen, es auf dem Master-Knoten speichern und es dann mit dem Befehl `sudo docker exec jupyterhub script_name` im Container `jupyterhub` ausführen. Das folgende Beispielskript installiert den Kernel und installiert dann einige Bibliotheken für diesen Kernel auf dem Master-Knoten, sodass Sie die Bibliotheken später unter Verwendung des Kernels in Jupyter importieren können.

```
#!/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
```

Um den Kernel und Bibliotheken innerhalb des Containers zu installieren, öffnen Sie eine Terminal-Verbindung zu dem Master-Knoten, speichern das Skript in `/etc/jupyter/install_kernels.sh` und führen den folgenden Befehl in der Master-Knoten Befehlszeile aus:

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

## Verwenden von Bibliotheken und Installieren von zusätzlichen Bibliotheken

Ein Kernsatz von Bibliotheken für maschinelles Lernen und Datenwissenschaft für Python 3 ist bei JupyterHub Amazon EMR vorinstalliert. Sie können `sudo docker exec jupyterhub bash -c "conda list"` und `sudo docker exec jupyterhub bash -c "pip freeze"` verwenden.

Wenn ein Spark-Job Bibliotheken auf Worker-Knoten benötigt, empfehlen wir, dass Sie beim Erstellen des Clusters ein Skript ausführen, um die Bibliotheken zu installieren. Bootstrap-Aktionen werden beim Erstellungsprozess des Clusters auf allen Cluster-Knoten ausgeführt, was die Installation vereinfacht. Wenn Sie Bibliotheken auf Core-/Worker-Knoten installieren, nachdem ein Cluster läuft, ist die Operation komplizierter. Wir zeigen in diesem Abschnitt ein Python-Beispielprogramm, das demonstriert, wie diese Bibliotheken zu installieren sind.

Die in diesem Abschnitt gezeigten Beispiele für Bootstrap-Aktionen und Python-Programme verwenden ein Bash-Skript, das in Amazon S3 gespeichert ist, um die Bibliotheken auf allen Knoten zu installieren.

Das in den folgenden Beispielen verwendete Skript `pip` installiert `paramiko`, `nlTK`, `scipy`, `scikit-learn` und `pandas` für den Python 3-Kernel:

```
#!/bin/bash


sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

Nachdem Sie das Skript erstellen, laden Sie es an einem Speicherort in Amazon S3 hoch, z. B. `s3://mybucket/install-my-jupyter-libraries.sh`. Weitere Informationen finden Sie unter [Hochladen von Objekten](#) im Benutzerhandbuch für Amazon Simple Storage Service, sodass Sie es in Ihrer Bootstrap-Aktion oder in Ihrem Python-Programm verwenden können.

Um eine Bootstrap-Aktion anzugeben, mit der Bibliotheken auf allen Knoten installiert werden, wenn Sie einen Cluster mit dem AWS CLI

1. Erstellen Sie ein Skript ähnlich dem früheren Beispiel und speichern Sie es an einem Speicherort in Amazon S3. Wir verwenden das Beispiel `s3://mybucket/install-my-jupyter-libraries.sh`.

- Erstellen Sie den Cluster mit JupyterHub und verwenden Sie das Path Argument der `--bootstrap-actions` Option, um den Speicherort des Skripts anzugeben, wie im folgenden Beispiel gezeigt:

 Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

So geben Sie eine Bootstrap-Aktion an, die Bibliotheken auf allen Knoten installiert, wenn Sie einen Cluster mithilfe der Konsole erstellen

- Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
- Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
- Geben Sie Einstellungen für Software and Steps (Software und Schritte) und Hardware an, wie für Ihre Anwendung geeignet.
- Auf dem Bildschirm General Cluster Settings (Allgemeine Cluster-Einstellungen) expandieren Sie Bootstrap Actions (Bootstrap-Aktionen).
- Wählen Sie für Add bootstrap action (Bootstrap-Aktion hinzufügen) die Option Custom action (Benutzerdefinierte Aktion) und Configure and add (Konfigurieren und hinzufügen) aus.
- Geben Sie für Name einen benutzerfreundlichen Namen ein. Geben Sie unter Skriptspeicherort den Speicherort Ihres Skripts in Amazon S3 ein (das Beispiel, das wir verwenden, ist `s3://mybucket/install-my-jupyter-libraries.sh`). Tragen Sie in Optional arguments (Optionale Argumente) nichts ein und wählen Sie Add (Hinzufügen).

7. Geben Sie andere Einstellungen für Ihren Cluster an und wählen Sie Next (Weiter).
8. Geben Sie Sicherheitseinstellungen an und wählen Sie Create cluster (Cluster erstellen).

### Example Installieren von Bibliotheken auf Core-Knoten eines aktiven Clusters

Nach der Installation von Bibliotheken auf dem Master-Knoten innerhalb von Jupyter können Sie auf verschiedene Arten Bibliotheken auf in Ausführung befindlichen Core-Knoten installieren. Das folgende Beispiel zeigt ein Python-Programm, das für die Ausführung auf einem lokalen Computer geschrieben wurde. Wenn Sie das Python-Programm lokal ausführen, verwendet es das AWS-RunShellScript von, AWS Systems Manager um das weiter oben in diesem Abschnitt gezeigte Beispielskript auszuführen, das Bibliotheken auf den Kernknoten des Clusters installiert.

```
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 Versionsverlauf

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon JupyterHub enthalten sindEMR, zusammen mit den Komponenten, die mit der Anwendung installiert

wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

### JupyterHub Versionsinformationen

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit JupyterHub                                                                                                                                                                                                                                                                                                            |
|----------------------------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.2.0                  | 1.5.0                 | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub                          |
| emr-5.36,2                 | 1.4.1                 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub |
| emr-7.1.0                  | 1.5.0                 | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-                                                                                                                                                                                                                                                                                                      |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit JupyterHub                                                                                                                                                                                                                                                                                   |
|----------------------------|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                            |                       | 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |



| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |



| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |



| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |



| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |

| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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 |



| Amazon EMR Release-Etikett | JupyterHub Ausführung | Komponenten, die installiert wurden mit 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

Livy ermöglicht die Interaktion über eine REST Schnittstelle mit einem EMR Cluster, auf dem Spark ausgeführt wird. Sie können die REST Schnittstelle oder eine RPC Client-Bibliothek verwenden, um Spark-Jobs oder Spark-Codefragmente einzureichen, Ergebnisse synchron oder asynchron abzurufen und Spark-Kontext zu verwalten. Weitere Informationen finden Sie auf der [Apache Livy-Website](#). Livy ist in der EMR Amazon-Release-Version 5.9.0 und höher enthalten.

Um auf die Livy-Weboberfläche zuzugreifen, richten Sie einen SSH Tunnel zum Master-Knoten und eine Proxyverbindung ein. Weitere Informationen finden Sie unter [Auf EMR Clustern gehostete Weboberflächen](#) anzeigen.

In der folgenden Tabelle sind die Version von Livy aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Livy EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Livy installiert wurden, finden Sie unter Komponentenversionen von Version 7.2.0.](#)

Livy-Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|---------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

In der folgenden Tabelle sind die Version von Livy aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Livy EMR installiert.

Die Version der Komponenten, die mit Livy in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Livy-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|---------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

In der folgenden Tabelle sind die Version von Livy aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Livy EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Livy installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Livy-Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                |
|---------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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- |

| EMRAmazon-Freigabeetikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                        |
|---------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                           |              | 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 |

## Themen

- [Aktivierung mit Apache Livy HTTPS](#)
- [Livy-Versionsverlauf](#)

## Aktivierung mit Apache Livy HTTPS

1. Stellen Sie einen EMR Amazon-Cluster mit aktivierter Übertragungsverschlüsselung bereit. Weitere Informationen zur Verschlüsselung finden Sie unter [Verschlüsseln von Daten im Ruhezustand und bei der Übertragung](#).
2. Erstellen Sie eine Datei mit dem Namen `livy_ssl.sh` und den folgenden Inhalten.

```
#!/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
```

- Führen Sie das folgende Skript als EMR Amazon-Schritt aus. Dieses Skript wird geändert, /etc/livy/conf/livy.conf um es zu aktivieren 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"}]'
```

- Starten Sie den Apache-Livy-Service neu, damit die Änderungen wirksam werden. Informationen zum Neustarten von Apache Livy finden Sie unter [Stoppen und Neustarten von Prozessen](#).
- Testen Sie, ob die Clients jetzt kommunizieren können mit HTTPS. Um beispielsweise einen Job einzureichen, führen Sie den folgenden Code aus.

```
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
```

Wenn Sie die Aktivierung HTTPS erfolgreich durchgeführt haben, sendet Livy eine Antwort, in der angegeben wird, dass der Befehl akzeptiert und der Batch-Job gesendet wurde.

```
{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo":{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ", "\nstderr: ", "\nYARN Diagnostics: "]}
```

## Livy-Versionsverlauf

In der folgenden Tabelle sind die Versionen von Livy aufgeführt, die in jeder Release-Version von Amazon enthalten sind EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

## Die Livy Versionsinformationen

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-                                                                                                                                                                                         |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                            |              | 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |



| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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                   |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |



| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |



| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                                        |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |



| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                                      |
|----------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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        |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|----------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|----------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|----------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|----------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |



| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|----------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                               |
|----------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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    |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                            |
|----------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                            |
|----------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                            |
|----------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                                            |
|----------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                   |
|----------------------------|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

| Amazon EMR Release-Etikett | Livy-Version | Mit Livy installierte Komponenten                                                                                                                                                                                                                                                                   |
|----------------------------|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.9.1                  | 0,4,0        | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server |
| emr-5.9.0                  | 0,4,0        | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server |



# Apache MXNet

Apache MXNet ist eine Beschleunigungsbibliothek, die für den Aufbau neuronaler Netzwerke und anderer Deep-Learning-Anwendungen entwickelt wurde. MXNet automatisiert gängige Arbeitsabläufe und optimiert numerische Berechnungen. MXNet hilft Ihnen dabei, neuronale Netzwerkarchitekturen zu entwerfen, ohne sich auf die Implementierung von Berechnungen auf niedriger Ebene, wie z. B. lineare Algebraoperationen, konzentrieren zu müssen. MXNet ist in der EMR Amazon-Release-Version 5.10.0 und höher enthalten.

Weitere Informationen finden Sie auf der [MXNet Apache-Website](#).

In der folgenden Tabelle sind die Version von MXNet aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. MXNet

Informationen zur Version der Komponenten, mit denen MXNet in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

MXNet Versionsinformationen für emr-7.2.0

| EMR Amazon-Freigabeetikett | MXNet Ausführung | Mit MXNet installierte Komponenten                                                                                                                                                                                                            |
|----------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

In der folgenden Tabelle sind die Version von MXNet aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. MXNet

Informationen zur Version der Komponenten, mit denen MXNet in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 6.15.0](#).

#### MXNetVersionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | MXNetAusführung | Mit MXNet installierte Komponenten                                                                                                                                                                                                            |
|---------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |

In der folgenden Tabelle sind die Version von MXNet aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. MXNet

Informationen zur Version der Komponenten, mit denen MXNet in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

#### MXNetVersionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | MXNetAusführung | Mit MXNet installierte Komponenten                                                                                                                                                       |
|---------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resour |

| EMR Amazon-Freigabeetikett | MXNet Ausführung | Mit MXNet installierte Komponenten                    |
|----------------------------|------------------|-------------------------------------------------------|
|                            |                  | cemanager, hadoop-yarn-timeline-server, mxnet, opencv |

## MXNet Verlauf der Veröffentlichung

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon MXNet enthalten sind EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

### MXNet Versionsinformationen

| Amazon EMR Release-Etikett | MXNet Ausführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.2.0                  | 1.9.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.36,2                 | 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-resour                                                      |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                            |                 | cemanager, 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-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 |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.14.0                 | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.13.0                 | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.12.0                 | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.0                 | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.9.1                  | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.9.0                  | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.1                  | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.8.0                  | 1.9.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.7.0                  | 1.8.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |



| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.1                 | 1.8.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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 |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.5.0                  | 1.8.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.4.0                  | 1.8.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.3.1                  | 1.7.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.3.0                  | 1.7.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.0.1                  | 1.5.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-6.0.0                  | 1.5.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.0                 | 1.7.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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 |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.2                 | 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-5.30.1                 | 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |



| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.29.0                 | 1.5.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.28.1                 | 1.5.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.28.0                 | 1.5.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.27.1                 | 1.4.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.27.0                 | 1.4.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.26.0                 | 1.4.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.25.0                 | 1.4.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.24.1                 | 1.4.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.24.0                 | 1.4.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.23.1                 | 1.3.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.2                 | 1.3.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.1                 | 1.3.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| 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 |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                      |
|----------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.17.0                 | 1.2.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |



| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.16.1                 | 1.2.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.16.0                 | 1.2.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.15.1                 | 1.1.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.15.0                 | 1.1.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.14.2                 | 1.1.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.14.1                 | 1.1.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                                       |
|----------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.14.0                 | 1.1.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv |
| emr-5.13.1                 | 1.0.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet         |
| emr-5.13.0                 | 1.0.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet         |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                               |
|----------------------------|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.3                 | 1.0.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet |
| emr-5.12.2                 | 1.0.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet |
| emr-5.12.1                 | 1.0.0           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit 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-resource-manager, hadoop-yarn-timeline-server, mxnet |
| emr-5.11.4                 | 0.12.0          | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet |
| emr-5.11.3                 | 0.12.0          | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                              |
|----------------------------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.2                 | 0.12.0          | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet |
| 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet |

| Amazon EMR Release-Etikett | MXNetAusführung | Komponenten wurden installiert mit MXNet                                                                                                                                                                                               |
|----------------------------|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.10.1                 | 0.12.0          | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, 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-resource-manager, hadoop-yarn-timeline-server, mxnet |

# Apache Oozie

Verwenden Sie den Apache Oozie Workflow Scheduler, um Hadoop-Aufträge zu verwalten und zu koordinieren. Weitere Informationen finden Sie unter <http://oozie.apache.org/>.

Die native Oozie-Weboberfläche wird von Amazon EMR nicht unterstützt. Wenn Sie eine Front-End-Oberfläche für Oozie nutzen möchten, können Sie die Hue Oozie-Anwendung verwenden. Weitere Informationen finden Sie unter [Hue](#). Oozie ist in der EMR Amazon-Release-Version 5.0.0 und höher enthalten. Oozie ist als Sandbox-Anwendung in früheren Versionen enthalten. Weitere Informationen finden Sie unter [Amazon EMR 4.x-Release-Versionen](#).

Wenn Sie ein benutzerdefiniertes Amazon Linux verwenden, das auf einem Amazon Linux AMI mit einem Erstellungsdatum vom 11.08.2018 AMI basiert, kann der Oozie-Server nicht gestartet werden. Wenn Sie Oozie verwenden, erstellen Sie eine benutzerdefinierte Version, die auf einer Amazon AMI Linux-ID mit einem anderen Erstellungsdatum AMI basiert. Sie können den folgenden AWS CLI Befehl verwenden, um eine Liste mit Images IDs für alle HVM Amazon Linux-Versionen AMIs mit einer Version 2018.03 zusammen mit dem Veröffentlichungsdatum zurückzugeben, sodass Sie ein geeignetes Amazon Linux AMI als Basis auswählen können. MyRegion Ersetzen Sie es durch Ihre Regionskennung, z. B. 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
```

In der folgenden Tabelle sind die Version von Oozie aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Oozie EMR installiert.

[Die Version der Komponenten, die in dieser Version mit Oozie installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Oozie-Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Oozie-Version | Mit Oozie installierte Komponenten                     |
|---------------------------|---------------|--------------------------------------------------------|
| emr-7.2.0                 | Oozie 5.2.1   | emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist- |



| EMRAmazon-Freigabeetikett | Oozie-Version | Mit Oozie installierte Komponenten                                                                                                                                                                                                                                                    |
|---------------------------|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                           |               | cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker |

In der folgenden Tabelle sind die Version von Oozie aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Oozie EMR installiert.

Die Version der Komponenten, die mit Oozie in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Oozie-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Oozie-Version | Mit Oozie installierte Komponenten                                                                                                                                                                                                                                                                |
|---------------------------|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                | Oozie 5.2.1   | emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, |

| EMRAmazon-Freigabeetikett | Oozie-Version | Mit Oozie installierte Komponenten       |
|---------------------------|---------------|------------------------------------------|
|                           |               | oozie-server, tez-on-yarn, tez-on-worker |

In der folgenden Tabelle sind die Version von Oozie aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Oozie EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Oozie installiert wurden, finden Sie unter Komponentenversionen der Version 5.36.2.](#)

Oozie-Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | Oozie-Version | Mit Oozie installierte Komponenten                                                                                                                                                                                                                                                                                         |
|---------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn |

## Themen

- [Verwenden von Oozie mit einer Remote-Datenbank in Amazon RDS](#)
- [Die Java-Version für Oozie konfigurieren](#)
- [Versionsverlauf von Oozie](#)

# Verwenden von Oozie mit einer Remote-Datenbank in Amazon RDS

Standardmäßig werden Oozie-Benutzerinformationen und Abfrageverläufe in einer lokalen SQL My-Datenbank auf dem Master-Knoten gespeichert. Alternativ können Sie einen oder mehrere Oozie-fähige Cluster mithilfe einer in Amazon S3 gespeicherten Konfiguration und einer Meine SQL Datenbank in Amazon Relational Database Service (Amazon) erstellen. RDS Auf diese Weise können Sie Benutzerinformationen und den von Oozie erstellten Abfrageverlauf beibehalten, ohne Ihren EMR Amazon-Cluster am Laufen zu halten. Wir empfehlen die Verwendung der serverseitigen Amazon-S3-Verschlüsselung zum Speichern der Konfigurationsdatei.

Erstellen Sie zunächst die Remote-Datenbank für Oozie.

Um die externe Datenbank Meine Datenbank zu erstellen SQL

1. Öffnen Sie die RDS Amazon-Konsole unter <https://console.aws.amazon.com/rds/>.
2. Klicken Sie auf Launch a DB Instance (DB-Instance starten).
3. Wählen Sie „Mein“ SQL und dann „Auswählen“.
4. Behalten Sie die Standardauswahl für Multi-AZ Deployment und Provisioned IOPS Storage bei und wählen Sie Weiter.
5. Übernehmen Sie die Standardwerte für die Instance-Spezifikationen, geben Sie Einstellungen an und wählen Sie Next (Weiter).
6. Wählen Sie auf der Seite „Configure Advanced Settings (Erweiterte Einstellungen konfigurieren)“ geeignete Sicherheitsgruppen- und Datenbanknamen aus. Die von Ihnen verwendete Sicherheitsgruppe muss mindestens eingehenden TCP Zugriff auf Port 3306 vom Master-Knoten Ihres Clusters aus zulassen. Wenn Sie zu diesem Zeitpunkt noch keinen Cluster erstellt haben, können Sie für alle Hosts eine Verbindung mit Port 3306 zulassen und die Sicherheitsgruppe anpassen, nachdem Sie den Cluster gestartet haben. Klicken Sie auf DB-Instance starten.
7. Wählen Sie im RDS Dashboard Instances und dann die Instanz aus, die Sie gerade erstellt haben. Wenn Ihre Datenbank verfügbar ist, notieren Sie sich den Datenbanknamen, den Benutzernamen, das Passwort und den RDS Instanz-Hostnamen. Diese Informationen benötigen Sie zum Erstellen und Konfigurieren Ihres Clusters.

Um eine externe MySQL Datenbank für Oozie anzugeben, wenn Sie einen Cluster starten, verwenden Sie AWS CLI

Um beim Starten eines Clusters mithilfe von einer externen MySQL-Datenbank für Oozie anzugeben AWS CLI, verwenden Sie die Informationen, die Sie bei der Erstellung Ihrer RDS-Instanz für die Konfiguration `oozie-site` mit einem Konfigurationsobjekt notiert haben.

**Note**

Sie können mehrere Cluster mit derselben externen Datenbank erstellen, aber jeder Cluster gibt Abfrageverläufe und Benutzerdaten frei.

- Erstellen Sie mit dem einen Cluster AWS CLI, auf dem Oozie installiert ist, verwenden Sie die externe Datenbank, die Sie erstellt haben, und verweisen Sie auf eine Konfigurationsdatei mit einer Konfigurationsklassifizierung für Oozie, die die Datenbankigenschaften spezifiziert. Im folgenden Beispiel wird ein Cluster mit installiertem Oozie erstellt, der auf die Konfigurationsdatei `myConfig.json` in Amazon S3 verweist, die die Datenbankkonfiguration angibt.

**Note**

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Beispielinhalte der `myConfig.json`-Datei werden unten angezeigt. Ersetzen *JDBC URL*, *username*, und *password* mit dem JDBC URL Benutzernamen und dem Passwort Ihrer Instanz. RDS

**⚠ Important**

Das JDBC URL muss den Datenbanknamen als Suffix enthalten. Zum Beispiel `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": []
}]
```

## Die Java-Version für Oozie konfigurieren

Oozie führt mehrere Java Virtual Machine () -Prozesse aus. JVM Auf dieser Seite wird erklärt, wie die Java-Version für jeden Prozess konfiguriert wird.

- Oozie-Server: Legen Sie mit `JAVA_HOME` in der `oozie-env`-Klassifizierung fest, dass die Java-Version für `EmbeddedOozieServer` aktualisiert werden soll.
- Oozie Launcher AM: Oozie Launcher AM ist ein Single-Mapper-MR-Job, der die entsprechenden Anwendungsbibliotheken wie Hadoop und Hive aufruft. Sofern nicht anders konfiguriert, entsprechen die Laufzeitversionen für Oozie Launcher AM den Java-Laufzeiten für Hadoop im Cluster. EMR Um die Java-Laufzeit für Oozie Launcher AM zu konfigurieren, legen Sie für den Job die folgende `workflow.xml` Eigenschaft fest:

```
<property>
 <name>mapred.child.env</name>
 <value>JAVA_HOME=/path/to/JAVA_HOME</value>
</property>
```

Diese Eigenschaft stellt sicher, dass Oozie Launcher AM für den Oozie-Auftrag auf der von Ihnen angegebenen Java-Version ausgeführt wird und nicht auf der Java-Version, die in Hadoop festgelegt ist.

- Ausführbare Datei des Anwendungsclients: Da Oozie Launcher AM den Anwendungsclient standardmäßig aufruft, ist die Java-Laufzeit für die ausführbare Client-Datei dieselbe wie für Oozie Launcher AM.
- Von einem Oozie-Job gestartete Anwendungen: Sofern nicht anders angegeben, entsprechen die Laufzeitversionen für die eigentliche Anwendung JVMs, die durch einen Oozie-Job gestartet werden, den Java-Laufzeiten für Hadoop im Cluster. EMR Abhängig von der Art der Oozie-Workflow-Aktion, die verwendet wird, um die Anwendung in einem Oozie-Job zu starten (Spark- oder Hive-Aktion), können Sie die Standard-Java-Laufzeit für die eigentliche Anwendung JVMs im für den Oozie-Job aktualisieren. `workflow.xml`

## Versionsverlauf von Oozie

In der folgenden Tabelle sind die in jeder Release-Version von Amazon EMR enthaltenen Versionen von Oozie sowie die mit der Anwendung installierten Komponenten aufgeführt. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

### Oozie-Versionsinformationen

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
		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-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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.15.0	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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.13.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.12.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.10.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.10.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
emr-5.36.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.6.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
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



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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



Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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

Amazon EMR Release-Etikett	Oozie-Version	Mit Oozie installierte Komponenten
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
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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

## Versionshinweise zu Oozie nach Version

### Themen

- [Amazon EMR 7.1.0 — Änderungen an Oozie](#)

## Amazon EMR 7.1.0 — Änderungen an Oozie

Typ	Beschreibung
Upgrade	Führt ein Upgrade von Netty auf 4.1.100.Final durch.
Upgrade	Führt ein Upgrade von Jetty auf 9.4.53.v2 0231009 durch.



# Apache Phoenix

Apache Phoenix wird für OLTP Betriebsanalysen verwendet, sodass Sie SQL Standardabfragen verwenden und JDBC APIs mit einem HBase Apache-Backing-Store arbeiten können. Weitere Informationen finden Sie unter [Phoenix in 15 minutes or less](#). Phoenix ist in der EMR Amazon-Release-Version 4.7.0 und höher enthalten.

Wenn Sie von einer früheren Version von Amazon EMR auf EMR Amazon-Release-Version 5.4.0 oder höher aktualisieren und die sekundäre Indizierung verwenden, aktualisieren Sie lokale Indizes wie in der [Apache Phoenix](#)-Dokumentation beschrieben. Amazon EMR entfernt die erforderlichen Konfigurationen aus der `hbase-site` Klassifizierung, aber die Indizes müssen neu gefüllt werden. Online- und Offline-Upgrades von Indizes werden unterstützt. Online-Upgrades sind die Standardeinstellung. Dies bedeutet, dass Indizes neu gefüllt werden, während die Initialisierung von Phoenix-Clients Version 4.8.0 oder höher ausgeführt wird. Um Offline-Upgrades anzugeben, setzen Sie die `phoenix.client.localIndexUpgrade` Konfiguration in der `phoenix-site` Klassifizierung auf `False` und dann SSH auf den Master-Knoten, der ausgeführt werden soll. `psql [zookeeper] -1`

In der folgenden Tabelle sind die Version von Phoenix aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Phoenix EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Phoenix installiert wurden, finden Sie unter [Komponentenversionen der Version 7.2.0](#).

Phoenix-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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-nodema

EMRAmazon-Freigabeetikett	Phoenix-Version	Mit Phoenix installierte Komponenten
		nager, 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

In der folgenden Tabelle sind die Version von Phoenix aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Phoenix EMR installiert.

Die Version der Komponenten, die mit Phoenix in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Phoenix-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.15.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-too

EMRAmazon-Freigabeetikett	Phoenix-Version	Mit Phoenix installierte Komponenten
		ls, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

In der folgenden Tabelle sind die Version von Phoenix aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Phoenix EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Phoenix installiert wurden, finden Sie unter Komponentenversionen der [Version 5.36.2](#).

#### Phoenix-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

## Themen

- [Einen Cluster mit Phoenix erstellen](#)
- [Phoenix-Clients](#)
- [Versionsverlauf von Phoenix](#)

## Einen Cluster mit Phoenix erstellen

Sie installieren Phoenix, indem Sie die Anwendung beim Erstellen eines Clusters in der Konsole oder über die AWS CLI auswählen. Die folgenden Verfahren und Beispiele zeigen, wie Sie einen Cluster mit Phoenix erstellen. Weitere Informationen zum Erstellen von Clustern mithilfe der Konsole, einschließlich erweiterter Optionen, finden Sie unter [Planen und Konfigurieren von Clustern](#) im Amazon EMR Management Guide.

So starten Sie einen Cluster mit installierter Phoenix-Anwendung mithilfe von Quick-Optionen zum Erstellen eines Clusters in der Konsole

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create cluster (Cluster erstellen) aus, um Quick Create (Schnellerstellung) zu verwenden.
3. Wählen Sie bei Software Konfiguration die neueste Version für Ihre Anwendung aus. Phoenix wird nur als Option angezeigt, wenn Amazon Release Version emr-4.7.0 oder höher ausgewählt ist.
4. Wählen Sie für Anwendungen die zweite Option HBase: HBase **ver** mit Ganglien **ver**, Hadoop **ver**, Bienenstock **ver**, Farbton **ver**, Phoenix **ver**, und ZooKeeper **ver**.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Das folgende Beispiel startet einen Cluster mit installierter Phoenix-Anwendung mithilfe von Standard-Konfigurationseinstellungen.

Um einen Cluster mit Phoenix zu starten und HBase den AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl:

```
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
```

## Anpassen von Phoenix-Konfigurationen

Beim Erstellen eines Clusters konfigurieren Sie Phoenix, indem Sie Werte in `hbase-site.xml` mithilfe der `hbase-site`-Konfigurationsklassifizierung festlegen.

Weitere Informationen finden Sie unter [Konfiguration und Optimierung](#) in der Phoenix-Dokumentation.

Das folgende Beispiel zeigt, wie eine in Amazon S3 gespeicherte JSON Datei verwendet wird, um den Wert von `false` für die `phoenix.schema.dropMetaData` Eigenschaft anzugeben. Für eine einzelne Klassifizierung können mehrere Eigenschaften angegeben werden. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#). Der `create-cluster` Befehl verweist dann auf die JSON Datei als `--configurations` Parameter.

Der Inhalt der in `/mybucket/myfolder/myconfig.json` gespeicherten JSON Datei lautet wie folgt.

```
[
 {
 "Classification": "hbase-site",
 "Properties": {
 "phoenix.schema.dropMetaData": "false"
 }
 }
]
```

Der `create cluster` Befehl, der auf die Datei verweist, wird im folgenden Beispiel gezeigt. JSON

```
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**

Neukonfigurationsanfragen für alle Phoenix-Konfigurationsklassifizierungen werden nur in EMR Amazon-Version 5.23.0 und höher unterstützt und in EMR Amazon-Version 5.21.0 oder 5.22.0 nicht unterstützt. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Phoenix-Clients

Sie stellen eine Verbindung zu Phoenix entweder über einen JDBC Client her, der mit vollständigen Abhängigkeiten erstellt wurde, oder über den „Thin Client“, der den Phoenix Query Server verwendet und nur auf einem Master-Knoten eines Clusters ausgeführt werden kann (z. B. mithilfe eines SQL Clients, eines Steps, einer Befehlszeile, einer Portweiterleitung usw.). SSH Wenn Sie den „Fat“ JDBC -Client verwenden, muss er dennoch Zugriff auf alle Knoten des Clusters haben, da er sich direkt mit HBase Diensten verbindet. Der "Thin" Phoenix-Client muss nur auf den Phoenix-Abfrageserver über einen Standard-Port 8765 zugreifen. Es gibt mehrere [Skripts](#) in Phoenix, die diese Clients verwenden.

Verwenden Sie einen EMR Amazon-Schritt, um eine Abfrage mit Phoenix durchzuführen

Das folgende Verfahren stellt einen Snapshot wieder her HBase und verwendet diese Daten, um eine Phoenix-Abfrage auszuführen. Sie können dieses Beispiel erweitern oder ein neues Skript erstellen, das die Phoenix-Clients an Ihre Anforderungen anpasst.

1. Erstellen Sie mit dem folgenden Befehl einen Cluster, auf dem Phoenix installiert ist:

```
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. Erstellen Sie die folgenden Dateien und laden Sie sie in Amazon S3: hoch:

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

Sie müssen das folgende Beispiel nur COLUMN\_ENCODED\_BYTES=0 angeben, wenn Sie EMR Amazon-Versionen 5.26.0 und höher verwenden.

```
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";
```

Verwenden Sie den AWS CLI , um die Dateien an den S3-Bucket zu senden:

```
aws s3 cp copySnapshot.sh s3://myBucket/
aws s3 cp runQuery.sh s3://myBucket/
```

```
aws s3 cp phoenixQuery.sql s3://myBucket/
```

- Erstellen Sie eine Tabelle mithilfe des folgenden, an den Cluster gesendeten Schritts, den Sie in Schritt 1 erstellt haben:

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
```

- Verwenden Sie `script-runner.jar`, um das zuvor in Ihren S3-Bucket hochgeladene `copySnapshot.sh`-Skript auszuführen:

```
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"]
```

Dadurch wird ein MapReduce Job ausgeführt, um Ihre Snapshot-Daten in den Cluster HDFS zu kopieren.

- Stellen Sie den Snapshot, den Sie in den Cluster kopiert haben, mit dem folgenden Schritt wieder her:

restoreSnapshot.json

```
[
 {
 "Name": "restore",
 "Args": ["bash", "-c", "echo '$disable \"customer\"; restore_snapshot \"customer_snapshot1\"; enable \"customer\" | hbase shell"],
 }
]
```



```

 "Jar": "command-runner.jar",
 "ActionOnFailure": "CONTINUE",
 "Type": "CUSTOM_JAR"
 }
]

```

```

aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./restoreSnapshot.json

```

6. Verwenden Sie `script-runner.jar`, um das zuvor in Ihren S3-Bucket hochgeladene `runQuery.sh`-Skript auszuführen:

```

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"]

```

Die Abfrage wird ausgeführt und gibt das Ergebnis an `stdout` des Schritts zurück. Es kann einige Minuten dauern, bis dieser Schritt abgeschlossen ist.

7. Überprüfen Sie die Ergebnisse der Schritte in dem ProtokollURI, das Sie `stdout` bei der Erstellung des Clusters in Schritt 1 verwendet haben. Die Ergebnisse sollten wie folgt aussehen:

```

+-----+-----+
| 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 |
+-----+-----+

```

## Versionsverlauf von Phoenix

In der folgenden Tabelle sind die Versionen von Phoenix aufgeführt, die in jeder Release-Version von Amazon enthalten sind EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

### Versionsinformationen für Phoenix

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-7.2.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
		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-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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.17.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.16.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.5.3	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.5.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.4.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-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, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.3.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.3.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.2.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.2.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, 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-5.1.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, 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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.1.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.0.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-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, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.9.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.9.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.9.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, 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-4.8.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, 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

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.8.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
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-ftpserver, 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-ftpserver, 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



Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.8.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.7.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-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-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, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR Release-Etikett	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.7.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-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, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

# Apache Pig

Apache Pig ist eine Open-Source-Apache-Bibliothek, die auf Hadoop basiert und eine Skriptsprache bereitstellt, die Sie verwenden können, um große Datasets zu transformieren, ohne dass Sie komplexen Code in einer Computersprache einer niedrigeren Ebene schreiben müssen, wie etwa Java. Die Bibliothek verwendet SQL ähnliche Befehle, die in einer Sprache namens Pig Latin geschrieben sind, und konvertiert diese Befehle in Tez-Jobs, die auf gerichteten azyklischen Graphen (DAGs) oder Programmen basieren. MapReduce Pig funktioniert mit strukturierten und unstrukturierten Daten in verschiedenen Formaten. Weitere Informationen zu Pig finden Sie unter <http://pig.apache.org/>.

Sie können Pig-Befehle interaktiv oder im Batch-Modus ausführen. Um Pig interaktiv zu verwenden, stellen Sie eine SSH Verbindung zum Master-Knoten her und senden Sie Befehle mithilfe der Grunt-Shell. Um Pig im Batch-Modus zu nutzen, laden Sie Ihre Pig-Skripts in Amazon S3 hoch und senden diese dann als Cluster-Schritte. Weitere Informationen zum Einreichen von Arbeit an einen Cluster finden Sie unter [Arbeit an einen Cluster einreichen](#) im Amazon EMR Management Guide.

Wenn Sie Pig verwenden, um Ausgaben in eine HCatalog Tabelle in Amazon S3 zu schreiben, deaktivieren Sie Amazon EMR Direct Write, indem Sie die `mapred.output.direct.EmrFileSystem` Eigenschaften `mapred.output.direct.NativeS3FileSystem` und auf `setzenfalse`. Weitere Informationen finden Sie unter [Verwenden HCatalog](#). Innerhalb eines Pig-Skripts können Sie die Befehle `SET mapred.output.direct.NativeS3FileSystem false` und `SET mapred.output.direct.EmrFileSystem false` verwenden.

In der folgenden Tabelle sind die Version von Pig aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Pig EMR installiert.

Die Version der Komponenten, die in dieser Version zusammen mit Pig installiert wurden, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

Pig-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Pig-Version	Mit Pig installierte Komponenten
emr-7.2.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-

EMR Amazon-Freigabeetikett	Pig-Version	Mit Pig installierte Komponenten
		cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

In der folgenden Tabelle sind die Version von Pig aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Pig EMR installiert.

Die Version der Komponenten, die mit Pig in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Pig-Versionsinformationen für emr-6.15.0

EMR Amazon-Freigabeetikett	Pig-Version	Mit Pig installierte Komponenten
emr-6.15.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

In der folgenden Tabelle sind die Version von Pig aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Pig EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Pig installiert wurden, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

Pig-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Pig-Version	Mit Pig installierte Komponenten
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

Themen

- [Übermitteln von Pig-Aufträgen](#)
- [Benutzerdefinierte Funktionen in Pig aufrufen](#)
- [Pig-Versionsverlauf](#)

## Übermitteln von Pig-Aufträgen

In diesem Abschnitt wird das Einreichen von Pig-Arbeiten an einen EMR Amazon-Cluster demonstriert. Das folgende Beispiel generiert einen Bericht mit der Gesamtzahl der übertragenen Bytes, eine Liste der Top 50 IP-Adressen, eine Liste der externen Top 50-Referrer und die 50 beliebtesten Suchbegriffe bei Bing und Google. Das Pig-Skript befindet sich im Amazon-S3-Bucket `s3://elasticmapreduce/samples/pig-apache/do-reports2.pig`. Die Eingabedaten

befinden sich im Amazon-S3-Bucket `s3://elasticmapreduce/samples/pig-apache/input`. Die Ausgabe wird in einem Amazon-S3-Bucket gespeichert.

## Arbeiten von Pig über die EMR Amazon-Konsole einreichen

In diesem Beispiel wird beschrieben, wie Sie mit der EMR Amazon-Konsole einem Cluster einen Pig-Schritt hinzufügen.

So senden Sie einen Pig-Auftrag

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Cluster erstellen, um einen neuen Cluster mit Pig zu erstellen. Anweisungen zum Erstellen eines Clusters finden Sie unter [Planen und Konfigurieren eines EMR Amazon-Clusters](#).
3. Öffnen Sie ein Terminal und SSH gelangen Sie in den Master-Knoten Ihres Clusters, indem Sie den Schritten folgen, die unter [Connect dem Master-Knoten herstellen mithilfe](#) beschrieben sind. Sobald Sie das getan haben, führen Sie die folgenden Schritte aus.

```
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. Klicken Sie in der Konsole die Seite Cluster-Liste aus und klicken Sie den Namen des von Ihnen erstellten Clusters.
5. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
6. Gehen Sie im Dialogfeld Schritt hinzufügen folgendermaßen vor:
  - Wählen Sie unter Step type (Schritttyp) die Option Pig program (Pig-Programm) aus.
  - Übernehmen Sie unter Name den Standardnamen („Pig program“) oder geben Sie einen neuen Namen ein.
  - Geben Sie unter Script S3 location (S3-Skriptspeicherort) den Speicherort Ihres Pig-Skripts ein. Zum Beispiel: **s3://elasticmapreduce/samples/pig-apache/doreports2.pig**.
  - Unter Input S3 location (S3-Eingabespeicherort) geben Sie den Speicherort der Eingabedaten ein. Zum Beispiel: **s3://elasticmapreduce/samples/pig-apache/input**.
  - Unter S3-Ausgabespeicherort geben Sie den Namen Ihres Amazon-S3-Ausgabe-Buckets ein oder navigieren zu dem entsprechenden S3-Bucket.

- Lassen Sie unter Arguments (Parameter) das Feld leer.
  - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
7. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
  8. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie über der Spalte Actions (Aktionen) das Symbol Refresh (Ansicht aktualisieren) aus, um den Status zu aktualisieren. Wenn Ihr Schritt abgeschlossen ist, überprüfen Sie Ihren Amazon-S3-Bucket, um sicherzustellen, dass die Ausgabedateien Ihres Pig-Schritts vorhanden sind.

## Reichen Sie Pig-Arbeiten ein, indem Sie AWS CLI

Um einen Pig-Schritt einzureichen, verwenden Sie AWS CLI

Wenn Sie einen Cluster mit dem starten AWS CLI, verwenden Sie den `--applications` Parameter, um Pig zu installieren. Verwenden Sie den Parameter `--steps`, um einen Pig-Schritt zu übermitteln.

1. Um einen Cluster zu starten, auf dem Pig installiert ist, geben Sie den folgenden Befehl ein und ersetzen *myKey* und *DOC-EXAMPLE-BUCKET/* mit dem Namen Ihres EC2 key pair und des Amazon S3 S3-Buckets.


```
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

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).



Wenn Sie die Instance-Anzahl ohne den Parameter `--instance-groups` angeben, wird ein einzelner Master-Knoten gestartet. Die verbleibenden Instances werden als Core-Knoten aufgerufen. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

 Note

Wenn Sie die EMR Standard-Servicerolle und das EC2 Instanzprofil noch nicht erstellt haben, geben Sie vor der Eingabe des `create-cluster` Unterbefehls ein, `aws emr create-default-roles` um sie zu erstellen.

- Um einen Pig-Schritt einzureichen, geben Sie den folgenden Befehl ein und ersetzen *myClusterId* und *DOC-EXAMPLE-BUCKET* mit Ihrer Cluster-ID und dem Namen Ihres Amazon S3 S3-Buckets.

```
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]
```

Dieser Befehl gibt eine Schritt-ID zurück, anhand derer Sie die Ausführung Ihres State-Schritts überprüfen können.

- Fragen Sie den Status Ihres Schritts mit dem `describe-step`-Befehl ab.

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

Der State-Wert des Schritts ändert sich mit der Ausführung des Schritts von PENDING zu RUNNING zu COMPLETED. Wenn Ihr Schritt abgeschlossen ist, überprüfen Sie Ihren Amazon-S3-Bucket, um sicherzustellen, dass die Ausgabedateien Ihres Pig-Schritts vorhanden sind.

Weitere Informationen zur Verwendung von EMR Amazon-Befehlen in der AWS CLI finden Sie in der [AWS CLI Befehlsreferenz](#).

## Benutzerdefinierte Funktionen in Pig aufrufen

Pig bietet die Möglichkeit, benutzerdefinierte Funktionen (UDFs) aus Pig-Skripten heraus aufzurufen. Sie können so eine benutzerdefinierte Verarbeitung in Ihren Pig-Skripten implementieren. Die derzeit unterstützten Sprachen sind Java, Python/Jython und JavaScript (obwohl die JavaScript Unterstützung noch experimentell ist.)

In den folgenden Abschnitten wird beschrieben, wie Sie Ihre Funktionen registrieren, um diese über die Pig-Shell oder in Pig-Skripten aufzurufen. Weitere Informationen zur Verwendung UDFs mit Pig finden Sie in der [Pig-Dokumentation für Ihre Version von Pig](#).

### Rufen Sie JAR Dateien von Pig aus auf

Sie können benutzerdefinierte JAR Dateien mit Pig verwenden, indem Sie den REGISTER Befehl in Ihrem Pig-Skript verwenden. Die JAR Datei ist lokal oder ein Remote-Dateisystem wie Amazon S3. Wenn das Pig-Skript ausgeführt wird, EMR lädt Amazon die JAR Datei automatisch auf den Master-Knoten herunter und lädt sie dann in den JAR verteilten Hadoop-Cache hoch. Auf diese Weise wird die JAR Datei bei Bedarf automatisch von allen Instances im Cluster verwendet.

Um JAR Dateien mit Pig zu verwenden

1. Laden Sie Ihre benutzerdefinierte JAR Datei in Amazon S3 hoch.
2. Verwenden Sie den REGISTER Befehl in Ihrem Pig-Skript, um den Bucket der benutzerdefinierten JAR Datei auf Amazon S3 anzugeben.

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

### Aufrufen von Python/Jython-Skripten in Pig

Sie können Python-Skripte in Pig registrieren und diese Skripte dann über die Pig-Shell oder in einem Pig-Skript aufrufen. Dies geschieht, indem Sie den Speicherort des Skripts mit dem Schlüsselwort `register` angeben.

Da Pig in Java geschrieben ist, verwendet es die Jython-Skript-Engine zum Parsen des Python-Skripts. Weitere Informationen zu Jython finden Sie unter <http://www.jython.org/>.

## So rufen Sie ein Python/Jython-Skript in Pig auf

1. Schreiben Sie ein Python-Skript und laden Sie es an einem Speicherort in Amazon S3 hoch. Es sollte sich um einen Bucket handeln, der sich im Besitz des Kontos befindet, das den Pig-Cluster erstellt bzw. das über entsprechende Berechtigungen verfügt. Nur so kann das Konto auf den Cluster zugreifen. In diesem Beispiel wird das Skript auf `s3://mybucket/pig/python` hochgeladen.
2. Starten Sie einen Pig-Cluster. Wenn Sie über die Grunt-Shell auf Pig zugreifen, starten Sie einen interaktiven Cluster. Wenn Sie Pig-Befehle aus einem Skript heraus ausführen, starten Sie einen skriptgesteuerten Pig-Cluster. In diesem Beispiel wird ein interaktiver Cluster gestartet. Weitere Informationen zum Erstellen eines Pig-Clusters finden Sie unter [Übermitteln von Pig-Aufträgen](#).
3. Verwenden Sie für einen interaktiven Cluster, SSH um eine Verbindung zum Master-Knoten herzustellen und die Grunt-Shell auszuführen. Weitere Informationen finden Sie unter [SSH into the Master Node](#).
4. Führen Sie die Grunt-Shell für Pig aus, indem Sie `pig` in der Befehlszeile eingeben:

```
pig
```

5. Registrieren Sie mit dem Schlüsselwort `register` in der Grunt-Eingabeaufforderung die Jython-Bibliothek und Ihr Python-Skript in Pig (wie im folgenden Beispiel gezeigt). Geben Sie den Speicherort Ihres Skripts in Amazon S3 an:

```
grunt> register 'lib/jython.jar';
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. Laden Sie die Eingabedaten. Beim folgenden Beispiel wird die Eingabe aus einem Amazon-S3-Speicherort geladen:

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as
(line:chararray);
```

7. Indem Sie mit `myfunctions` auf sie verweisen, können Sie jetzt Funktionen in Ihrem Pig-Skript aufrufen:

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

## Pig-Versionsverlauf

In der folgenden Tabelle sind die Version von Pig aufgeführt, die in jeder Release-Version von Amazon enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Pig Versionsinformationen

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-https-server, hadoop-kms-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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
emr-4.9.4	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
emr-4.8.5	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.8.4	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
emr-4.8.3	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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
emr-4.7.4	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.7.3	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client



Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-httplibs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
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

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
emr-4.4.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.3.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.2.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR Release-Etikett	Pig-Version	Mit Pig installierte Komponenten
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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.0.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

# Presto und Trino

## Note

Presto SQL wurde im Dezember 2020 in Trino umbenannt. EMRAmazon-Versionen 6.4.0 und höher verwenden den Namen Trino, während frühere Versionen den Namen Presto verwenden. SQL

[Presto](#) ist eine schnelle SQL Abfrage-Engine, die für interaktive Analyseabfragen über große Datensätze aus mehreren Quellen entwickelt wurde. Weitere Informationen finden Sie auf der [Presto-Website](#). Presto ist in den EMR Amazon-Versionen 5.0.0 und höher enthalten. Frühere Versionen enthalten Presto als Sandbox-Anwendung. Weitere Informationen finden Sie unter [Amazon EMR 4.x-Release-Versionen](#). EMRAmazon-Release-Versionen 6.1.0 und höher unterstützen zusätzlich zu Presto auch [Trino](#) (PrestoSQL). Weitere Informationen finden Sie unter [Installation von PrestoDB und Trino](#).

In der folgenden Tabelle sind die Version von Presto aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Presto EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Presto installiert wurden, finden Sie unter Komponentenversionen von Version 7.2.0.](#)

Presto-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Presto-Version	Mit Presto installierte Komponenten
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,

EMRAmazon-Freigabeetikett	Presto-Version	Mit Presto installierte Komponenten
		hcatalog-server, mariadb-server, presto-coordinator, presto-worker

In der folgenden Tabelle sind die Version von Presto aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Presto EMR installiert.

Die Version der Komponenten, die mit Presto in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Presto-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Presto-Version	Mit Presto installierte Komponenten
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

In der folgenden Tabelle sind die Version von Presto aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Presto EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Presto installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

## Presto-Versionsinformationen für emr-5.36.2

EMRAmazon-Freigabeetikett	Presto-Version	Mit Presto installierte Komponenten
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

In der folgenden Tabelle sind die Version von Trino (PrestoSQL) aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon mit Trino (Presto) EMR installiert. SQL

[Informationen zur Version der Komponenten, die mit Trino \(PrestoSQL\) in dieser Version installiert wurden, finden Sie unter Komponentenversionen der Version 6.15.0.](#)

## Versionsinformationen von Trino (SQLPresto) für emr-6.15.0

EMRAmazon-Freigabeetikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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,

EMRAmazon-Freigabeetikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
---------------------------	------------------------------	---------------------------------------------------------------------------------------------------------------------------------------

## Themen

- [Presto mit dem AWS Glue-Datenkatalog verwenden](#)
- [S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden](#)
- [Hinzufügen von Datenbank-Konnektoren](#)
- [SSLPresto auf Amazon verwenden TLS und LDAPS mit Presto konfigurieren EMR](#)
- [Den strikten Presto-Modus aktivieren](#)
- [Umgang mit dem Verlust von Spot Instances in Presto](#)
- [Fehlertolerante Ausführung in Trino](#)
- [Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden](#)
- [Überlegungen zu Presto bei Amazon EMR](#)
- [Presto-Versionsverlauf](#)

## Presto mit dem AWS Glue-Datenkatalog verwenden

Mit der EMR Amazon-Release-Version 5.10.0 und höher können Sie den AWS Glue-Datenkatalog als Standard-Hive-Metastore für Presto angeben. Wir empfehlen diese Konfiguration, wenn Sie einen persistenten Metastore oder einen gemeinsamen Metastore für verschiedene Cluster, Services, Anwendungen oder AWS-Konten benötigen.

AWS Glue ist ein vollständig verwalteter Service zum Extrahieren, Transformieren und Laden (ETL), mit dem Sie Ihre Daten einfach und kostengünstig kategorisieren, bereinigen, anreichern und zuverlässig zwischen verschiedenen Datenspeichern verschieben können. Der AWS Glue Data Catalog bietet ein einheitliches Metadaten-Repository für eine Vielzahl von Datenquellen und Datenformaten, das sich sowohl in Amazon EMR als auch in AmazonRDS, Amazon Redshift, Redshift Spectrum, Athena und jede mit dem Apache Hive Metastore kompatible Anwendung integrieren lässt. AWS Glue-Crawler können automatisch Schemas aus Quelldaten



in Amazon S3 ableiten und die zugehörigen Metadaten im Datenkatalog speichern. Weitere Informationen zum Datenkatalog finden Sie unter [Füllen des AWS Glue-Datenkatalogs im AWS Glue-Entwicklerhandbuch](#).

Für AWS Glue fallen separate Gebühren an. Es gibt eine monatliche Gebühr für das Speichern und Zugreifen auf die Metadaten im Datenkatalog, einen Stundensatz, der pro Minute für AWS ETL Glue-Jobs und Crawler-Laufzeit abgerechnet wird, und einen Stundensatz, der pro Minute für jeden bereitgestellten Entwicklungsendpunkt abgerechnet wird. Der Datenkatalog ermöglicht Ihnen die gebührenfreie Speicherung von bis zu einer Million Objekten. Wenn Sie mehr als eine Million Objekte speichern, wird Ihnen USD 1 USD für jeweils 100.000 Objekte über eine Million berechnet. Im Data Catalog ist ein Objekt eine Tabelle, eine Partition oder eine Datenbank. Weitere Informationen finden Sie unter [Glue-Preise](#).

#### Important

Wenn Sie vor dem 14. August 2017 Tabellen mit Amazon Athena oder Amazon Redshift Spectrum erstellt haben, werden Datenbanken und Tabellen in einem von Athena verwalteten Katalog gespeichert, der vom Glue-Datenkatalog getrennt ist. AWS Um Amazon EMR in diese Tabellen zu integrieren, müssen Sie ein Upgrade auf den AWS Glue Data Catalog durchführen. Weitere Informationen finden Sie unter [Upgrade auf den AWS Glue-Datenkatalog](#) im Amazon Athena Athena-Benutzerhandbuch.

## AWS Glue Data Catalog als Metastore angeben

Sie können den AWS Glue-Datenkatalog als Metastore angeben, indem Sie Amazon AWS Management Console EMR API, AWS CLI, oder verwenden. Wenn Sie das CLI oder verwenden API, verwenden Sie die Konfigurationsklassifizierung für Presto, um den Datenkatalog anzugeben. Darüber hinaus können Sie mit Amazon EMR 5.16.0 und höher die Konfigurationsklassifizierung verwenden, um einen Datenkatalog in einem anderen zu spezifizieren. AWS-Konto Wenn Sie die Konsole verwenden, können Sie den Data Catalog mit den erweiterten Optionen oder den Schnelloptionen angeben.

### Console

Um AWS Glue Data Catalog mit der neuen Konsole als Hive-Metastore anzugeben

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.

2. Wählen Sie EC2 im linken Navigationsbereich unter EMR on die Option Clusters und anschließend Create cluster aus.
3. Wählen Sie unter Anwendungspaket die Option Presto aus.
4. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Presto-Tabellenmetadaten verwenden.
5. Wählen Sie alle anderen Optionen aus, die für Ihren Cluster gelten.
6. Um Ihren Cluster jetzt zu starten, wählen Sie Cluster erstellen aus.

## CLI

Um den AWS Glue-Datenkatalog als Standard-Hive-Metastore anzugeben, verwenden Sie den AWS CLI

Beispiele dafür, wie Sie die folgenden Konfigurationsklassifizierungen unten angeben, wenn Sie einen Cluster erstellen, finden Sie unter [Anwendungen konfigurieren](#).

Amazon EMR 5.16.0 und höher

- Stellen Sie die `hive.metastore` Eigenschaft auf ein, `glue` wie im folgenden JSON Beispiel gezeigt.

```
[
 {
 "Classification": "presto-connector-hive",
 "Properties": {
 "hive.metastore": "glue"
 }
 }
]
```

Um einen Datenkatalog in einem anderen Format anzugeben AWS-Konto, fügen Sie die `hive.metastore.glue.catalogid` Eigenschaft wie im folgenden JSON Beispiel gezeigt hinzu. Ersetzen Sie *acct-id* durch das AWS -Konto des Data Catalog. Die Verwendung eines Datenkatalogs in einem anderen AWS-Konto ist mit EMR Amazon-Version 5.15.0 und früher nicht verfügbar.

```
[
 {
 "Classification": "presto-connector-hive",
```

```
"Properties": {
 "hive.metastore": "glue",
 "hive.metastore.glue.catalogid": "acct-id"
}
}
```

### Amazon EMR 5.10.0 bis 5.15.0

Stellen Sie die `hive.metastore.glue.datacatalog.enabled` Eigenschaft auf `true`, wie im folgenden Beispiel gezeigt: JSON

```
[
 {
 "Classification": "presto-connector-hive",
 "Properties": {
 "hive.metastore.glue.datacatalog.enabled": "true"
 }
 }
]
```

### Amazon EMR 6.1.0 und höher mit Presto SQL (Trino)

Ab EMR Version 6.1.0 unterstützt Presto SQL auch Glue als Standard-Hive-Metastore. Verwenden Sie die `prestoql-connector-hive` Konfigurationsklassifizierung und setzen Sie die `hive.metastore` Eigenschaft auf `glue`, wie im folgenden Beispiel gezeigt. JSON

EMRAmazon-Versionen 6.4.0 und höher verwenden den neuen Namen Trino anstelle von Presto. SQL Wenn Sie Trino verwenden, ersetzen Sie die Klassifizierung `prestoql-connector-hive` in der folgenden Konfiguration mit `trino-connector-hive` durch.

```
[
 {
 "Classification": "prestoql-connector-hive",
 "Properties": {
 "hive.metastore": "glue"
 }
 }
]
```

Um Metastores auf einem langlebigen Cluster zu wechseln, können Sie manuell die für die Version geeigneten Werte festlegen, indem Sie eine Verbindung zum Master-Knoten herstellen, die Eigenschaftswerte direkt in der Datei `/etc/presto/conf/catalog/hive.properties` bearbeiten und den Presto-Server (`sudo restart presto-server`) neu starten. Wenn Sie diese Methode mit Amazon EMR 5.15.0 und früher verwenden, stellen Sie sicher, dass sie auf eingestellt `hive.table-statistics-enabled` ist. `false` Diese Einstellung ist nicht erforderlich, wenn Sie Versionen ab 5.16.0 verwenden. Trotzdem werden Tabellen- und Partitionsstatistiken nicht unterstützt.

## IAMBerechtigungen

Das EC2 Instanzprofil für einen Cluster muss über IAM Berechtigungen für AWS Glue-Aktionen verfügen. Wenn Sie die Verschlüsselung für AWS Glue Data Catalog-Objekte aktivieren, muss die Rolle außerdem berechtigt sein, die für die Verschlüsselung AWS KMS key verwendeten Daten zu verschlüsseln, zu entschlüsseln und zu generieren.

### Berechtigungen für AWS Glue-Aktionen

Wenn Sie das EC2 Standard-Instance-Profil für Amazon verwenden, ist keine Aktion erforderlich. Die `AmazonElasticMapReduceforEC2Role` verwaltete Richtlinie, die an die angehängt ist, `EMR_EC2_DefaultRole` ermöglicht alle erforderlichen AWS Glue-Aktionen. Wenn Sie jedoch ein benutzerdefiniertes EC2 Instanzprofil und Berechtigungen angeben, müssen Sie die entsprechenden AWS Glue-Aktionen konfigurieren. Verwenden Sie hierzu die verwaltete Richtlinie `AmazonElasticMapReduceforEC2Role` als Ausgangspunkt. Weitere Informationen finden Sie unter [Servicerolle für EC2 Cluster-Instances \(EC2Instance-Profil\)](#) im Amazon EMR Management Guide.

### Berechtigungen zum Verschlüsseln und Entschlüsseln von AWS Glue Data Catalog

Ihr Instance-Profil benötigt die Erlaubnis, Daten mithilfe Ihres Schlüssels zu verschlüsseln und zu entschlüsseln. Sie müssen diese Berechtigungen nicht konfigurieren, wenn die beiden folgenden Aussagen zutreffen:

- Sie aktivieren die Verschlüsselung für AWS Glue Data Catalog-Objekte mithilfe von verwalteten Schlüsseln für AWS Glue.
- Sie verwenden einen Cluster, der sich im selben Verzeichnis befindet AWS-Konto wie der AWS Glue-Datenkatalog.

Andernfalls müssen Sie der Berechtigungsrichtlinie, die Ihrem EC2 Instanzprofil beigefügt ist, die folgende Erklärung hinzufügen.

```
[
 {
 "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"
 }
]
 }
]
```

Weitere Informationen zur Verschlüsselung von AWS Glue Data Catalog finden Sie unter [Verschlüsselung Ihres Datenkatalogs](#) im AWS Glue Developer Guide.

## Ressourcenbasierte Berechtigungen

Wenn Sie AWS Glue in Verbindung mit Hive, Spark oder Presto in Amazon EMR verwenden, unterstützt AWS Glue ressourcenbasierte Richtlinien zur Steuerung des Zugriffs auf Datenkatalogressourcen. Zu diesen Ressourcen gehören Datenbanken, Tabellen, Verbindungen und benutzerdefinierte Funktionen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für AWS Glue](#) im AWS -Glue-Entwicklerhandbuch.

Wenn Sie ressourcenbasierte Richtlinien verwenden, um den Zugriff auf AWS Glue von Amazon aus zu beschränken, muss der Principal, den Sie in der Berechtigungsrichtlinie angeben, die Rolle sein, die dem EC2 Instance-Profil ARN zugeordnet ist, das bei der Erstellung eines Clusters angegeben wird. Beispielsweise können Sie für eine ressourcenbasierte Richtlinie, die an einen Katalog angehängt ist, die Rolle ARN für die Standard-Service-Rolle für Cluster-Instances angeben. EC2 *EMR\_EC2\_DefaultRole* als der Principal, wobei das im folgenden Beispiel gezeigte Format verwendet wird:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Das Tool *acct-id* kann sich von der AWS Glue-Konto-ID unterscheiden. Dies ermöglicht den Zugriff von EMR Clustern in verschiedenen Konten aus. Sie können mehrere Principals angeben, von denen jeder aus einem anderen Konto stammt.

## Überlegungen zur Verwendung von AWS Glue Data Catalog

Beachten Sie die folgenden Punkte, wenn Sie AWS Glue Data Catalog als Metastore mit Presto verwenden:

- Das Umbenennen von Tabellen in AWS Glue wird nicht unterstützt.
- Wenn Sie eine Hive-Tabelle erstellen, ohne eine LOCATION anzugeben, werden die Tabellendaten an dem in der Eigenschaft `hive.metastore.warehouse.dir` angegebenen Ort gespeichert. Standardmäßig ist dies ein Speicherort in HDFS. Wenn ein anderer Cluster auf die Tabelle zugreifen muss, schlägt er fehl, sofern er nicht über ausreichende Berechtigungen für den Cluster verfügt, der die Tabelle erstellt hat. Da der HDFS Speicher vorübergehend ist, gehen außerdem die Tabellendaten verloren, wenn der Cluster beendet wird, und die Tabelle muss neu erstellt werden. Wir empfehlen, dass Sie LOCATION in Amazon S3 angeben, wenn Sie eine Hive-Tabelle mit AWS Glue erstellen. Alternativ können Sie die `hive-site`-Konfigurationsklassifizierung verwenden, um einen Speicherort in Amazon S3 für `hive.metastore.warehouse.dir` anzugeben, der für alle Hive-Tabellen gilt. Wenn eine Tabelle an einem HDFS Standort erstellt wurde und der Cluster, der sie erstellt hat, noch läuft, können Sie den Tabellenstandort von AWS Glue aus auf Amazon S3 aktualisieren. Weitere Informationen finden Sie unter [Arbeiten mit Tabellen auf der AWS Glue-Konsole](#) im AWS Glue-Entwicklerhandbuch.
- Partitionswerte, die Anführungszeichen und Apostrophe enthalten, werden nicht unterstützt, zum Beispiel `PARTITION (owner="Doe's")`.
- [Spaltenstatistiken](#) werden für emr-5.31.0 und höher unterstützt.
- Die Verwendung der [Hive-Autorisierung](#) wird nicht unterstützt. Als Alternative sollten Sie die Verwendung [ressourcenbasierter Richtlinien für AWS Glue](#) in Betracht ziehen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für den Amazon EMR Access to AWS Glue-Datenkatalog](#).

# S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden

Mit der EMR Amazon-Release-Version 5.18.0 und höher können Sie [S3 Select](#) Pushdown mit Presto bei Amazon verwenden. Dieses Feature ermöglicht Presto Rechenarbeit der Projektionsoperationen (z. B. SELECT) und der Prädikatoperationen (z. B. WHERE) in Amazon S3 herunter zu schieben („Pushdown“). Auf diese Weise können Abfragen nur die erforderlichen Daten von Amazon S3 abrufen, wodurch die Leistung verbessert und die zwischen Amazon EMR und Amazon S3 übertragene Datenmenge in einigen Anwendungen reduziert werden kann.

## Ist S3-Select-Pushdown das Richtige für meine Anwendung?

Wir empfehlen, dass Sie Benchmark-Tests für Ihre Anwendungen im Vergleich mit und ohne S3 Select-Pushdown durchführen, um zu sehen, ob es für Ihre Anwendung geeignet sein könnte.

Verwenden Sie die folgenden Richtlinien, um zu bestimmen, ob Ihre Anwendung ein Kandidat für die Verwendung von S3 ist:

- Ihre Abfrage filtert mehr als die Hälfte des ursprünglichen Datensatzes.
- Ihre Abfrage-Filterprädikate verwenden Spalten mit einem Datentyp, der von Presto und S3 Select unterstützt wird. Die Datentypen timestamp, real und double werden von S3 Select-Pushdown nicht unterstützt. Wir empfehlen für numerische Daten die Verwendung des Datentyps decimal. Weitere Informationen zu unterstützten Datentypen für S3 Select finden Sie unter [Datentypen](#) im Benutzerhandbuch für Amazon Simple Storage Service.
- Ihre Netzwerkverbindung zwischen Amazon S3 und dem EMR Amazon-Cluster hat eine gute Übertragungsgeschwindigkeit und verfügbare Bandbreite. Amazon S3 komprimiert HTTP Antworten nicht, daher wird die Antwortgröße bei komprimierten Eingabedateien wahrscheinlich zunehmen.

## Überlegungen und Einschränkungen

- Es werden nur Objekte unterstützt, die im CSV Format gespeichert sind. Objekte können optional mit gzip oder bzip2 komprimiert bzw. dekomprimiert werden.
- Die Eigenschaft AllowQuotedRecordDelimiters wird nicht unterstützt. Wenn diese Eigenschaft angegeben ist, schlägt die Abfrage fehl.

- Serverseitige Amazon S3 S3-Verschlüsselung mit vom Kunden bereitgestellten Verschlüsselungsschlüsseln (SSE-C) und clientseitige Verschlüsselung werden nicht unterstützt.
- S3 Select Pushdown ist kein Ersatz für die Verwendung von spaltenförmigen oder komprimierten Dateiformaten wie Parquet. ORC

## Aktivieren von S3-Select-Pushdown für Presto oder Trino

Um S3 Select Pushdown for PrestoDB auf Amazon zu aktivierenEMR, verwenden Sie die `presto-connector-hive` Konfigurationsklassifizierung, `hive.s3select-pushdown.enabled` auf eingestellt, `true` wie im folgenden Beispiel gezeigt. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#). Außerdem muss der Wert für „hive.s3select-pushdown.max-connections“ festgelegt werden. Für die meisten Anwendungen sollte die Standardeinstellung **500** ausreichen. Weitere Informationen hierzu finden Sie unter [Verstehen und Optimieren von „hive.s3select-pushdown.max-connections“](#) weiter unten.

Ersetzen *presto-connector-hive* Sie für Presto SQL auf den EMR Versionen 6.1.0 bis 6.3.0 im Beispiel unten durch. `prestoql-connector-hive`

EMR Amazon-Versionen 6.4.0 und höher verwenden den neuen Namen Trino anstelle von Presto. SQL Wenn Sie Trino verwenden, ersetzen Sie im Beispiel unten *presto-connector-hive* durch `trino-connector-hive`

```
[
 {
 "classification": "presto-connector-hive",
 "properties": {
 "hive.s3select-pushdown.enabled": "true",
 "hive.s3select-pushdown.max-connections": "500"
 }
 }
]
```

## Verstehen und Optimieren von „hive.s3select-pushdown.max-connections“

Standardmäßig verwendet Presto EMRFS als Dateisystem. Die Einstellung `fs.s3.maxConnections` in der `emrfs-site` Konfigurationsklassifizierung legt die maximal zulässigen Client-Verbindungen zu Amazon S3 EMRFS für Presto fest. Standardmäßig ist dies 500. S3 Select Pushdown umgeht EMRFS beim Zugriff auf Amazon S3 für Prädikatsoperationen.



In diesem Fall gibt der Wert von `hive.s3select-pushdown.max-connections` die maximale Anzahl von Clientverbindungen an, die für diese Operationen von Worker-Knoten aus zulässig ist. Für alle Anfragen an Amazon S3, die Presto initiiert und die nicht nach unten verschoben werden (z. B. GET Operationen), gilt jedoch weiterhin der Wert von `fs.s3.maxConnections`.

Wenn Ihre Anwendung den Fehler „Timeout waiting for connection from pool (Timeout beim Warten auf eine Verbindung aus dem Pool)“ erhält, erhöhen Sie den Wert von `hive.s3select-pushdown.max-connections` und von `fs.s3.maxConnections`.

## Hinzufügen von Datenbank-Konnektoren

Sie können Konfigurationsklassifizierungen verwenden, um Konnektoreigenschaften zu konfigurieren JDBC, wenn Sie einen Cluster erstellen. Konfigurationsklassifizierungen beginnen mit `presto-connector`, z. B. `presto-connector-postgresql`. Die verfügbaren Konfigurationsklassifizierungen hängen von der EMR Amazon-Release-Version ab. Die mit der neuesten Release-Version verfügbaren Konfigurationsklassifizierungen finden Sie unter Amazon [the section called “5.36.2 Konfigurationsklassifizierungen”](#) EMR 5.36.2. Wenn Sie eine andere Version von Amazon verwenden EMR, finden Sie [Amazon EMR 5.x-Release-Versionen](#) die Konfigurationsklassifizierungen unter. Weitere Informationen zu den Eigenschaften, die mit den einzelnen Konnektoren konfiguriert werden können, finden Sie unter <https://prestodb.io/docs/current/connector.html>.

Example — Konfiguration eines Clusters mit dem SQL JDBC Postgre-Connector

Um einen Cluster mit installiertem und konfiguriertem SQL Postgre-Connector zu starten, erstellen Sie zunächst eine JSON Datei, die die Konfigurationsklassifizierung spezifiziert, z. B. `myConfig.json` mit dem folgenden Inhalt, und speichern Sie sie lokal.

Ersetzen Sie die Verbindungseigenschaften entsprechend Ihrem Setup und wie im Thema [SQL Postgre-Konnektor](#) in der Presto-Dokumentation beschrieben.

```
[
 {
 "Classification": "presto-connector-postgresql",
 "Properties": {
 "connection-url": "jdbc:postgresql://example.net:5432/database",
 "connection-user": "MYUSER",
 "connection-password": "MYPASS"
 },
 "Configurations": []
 }
]
```

```
}
]
```

Wenn Sie den Cluster erstellen, referenzieren Sie den Pfad zur JSON Datei mithilfe der `--configurations` Option, wie im folgenden Beispiel gezeigt, wobei sich die Datei in demselben Verzeichnis `myConfig.json` befindet, in dem Sie den Befehl ausführen:

```
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
```

## SSL Presto auf Amazon verwenden TLS und LDAPS mit Presto konfigurieren EMR

Mit der EMR Amazon-Version 5.6.0 und höher können Sie SSL aktivieren, TLS um die [interne Kommunikation zwischen Presto-Knoten zu sichern](#). Zu diesem Zweck richten Sie eine Sicherheitskonfiguration für die Verschlüsselung der Daten während der Übertragung ein. Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#) und [Verwenden von Sicherheitskonfigurationen zur Einrichtung der Clustersicherheit](#) im Amazon EMR Management Guide.

Wenn Sie eine Sicherheitskonfiguration mit Verschlüsselung während der Übertragung verwenden, EMR geht Amazon für Presto wie folgt vor:

- Der Service verteilt die von Ihnen angegebenen Verschlüsselungsartefakte oder Zertifikate über den Presto-Cluster. Weitere Informationen finden Sie unter [Bereitstellen von Zertifikaten für die Datenverschlüsselung während der Übertragung](#).
- Legt die folgenden Eigenschaften mit der Konfigurationsklassifizierung `presto-config` fest, die der `config.properties`-Datei für Presto entspricht:
  - Legt `http-server.http.enabled false` auf allen Knoten fest, was zu Gunsten von deaktiviert wird HTTP. HTTPS Dazu müssen Sie bei der Einrichtung der Sicherheitskonfiguration für die Verschlüsselung DNS während der Übertragung Zertifikate angeben, die für öffentlich und privat funktionieren. Eine Möglichkeit, dies zu tun, besteht darin, Zertifikate SAN (Subject Alternative Name) zu verwenden, die mehrere Domänen unterstützen.

- Legt `http-server.https.*`-Werte fest. Einzelheiten zur Konfiguration finden Sie in der Presto-Dokumentation unter [LDAPAuthentifizierung](#).
- Für Presto SQL (Trino) auf EMR Version 6.1.0 und höher konfiguriert Amazon EMR automatisch einen gemeinsamen geheimen Schlüssel für die sichere interne Kommunikation zwischen Clusterknoten. Sie müssen keine zusätzliche Konfiguration vornehmen, um dieses Sicherheitsfeature zu aktivieren, und Sie können die Konfiguration mit Ihrem eigenen geheimen Schlüssel überschreiben. Informationen zur internen Authentifizierung von Trino finden Sie in der [Trino-353-Dokumentation: Sichere interne Kommunikation](#).

Darüber hinaus können Sie mit der EMR Amazon-Release-Version 5.10.0 und höher die [LDAPAuthentifizierung](#) für Client-Verbindungen zum Presto-Koordinator mithilfe von einrichten. HTTPS Dieses Setup verwendet secure LDAP (LDAPS). TLS muss auf Ihrem LDAP Server aktiviert sein und der Presto-Cluster muss eine Sicherheitskonfiguration mit aktivierter Datenverschlüsselung bei der Übertragung verwenden. Zusätzliche Konfigurationsschritte sind erforderlich. Die Konfigurationsoptionen unterscheiden sich je nach der von Ihnen verwendeten Release-Version von Amazon EMR. Weitere Informationen finden Sie unter [LDAPAuthentifizierung für Presto bei Amazon verwenden EMR](#).

Presto bei Amazon EMR verwendet standardmäßig Port 8446 für interne HTTPS Zwecke. Der für die interne Kommunikation verwendete Port muss derselbe Port sein, der für den HTTPS Client-Zugriff auf den Presto-Koordinator verwendet wird. Die Eigenschaft `http-server.https.port` in der Konfigurationsklassifizierung `presto-config` gibt den Port an.

## LDAPAuthentifizierung für Presto bei Amazon verwenden EMR

Folgen Sie zur Konfiguration den Schritten in diesem Abschnitt LDAP. Beachten Sie die einzelnen Schritte sowie die Beispiele und Links mit weiterführenden Informationen.

### Schritte zur Konfiguration der LDAP Authentifizierung

- [Schritt 1: Sammeln Sie Informationen über Ihren LDAP Server und kopieren Sie das Serverzertifikat auf Amazon S3](#)
- [Schritt 2: Sicherheitskonfiguration einrichten](#)
- [Schritt 3: Erstellen Sie eine Konfiguration JSON mit Presto-Eigenschaften für LDAP](#)
- [Schritt 4: Erstellen Sie das Skript, um das LDAP Serverzertifikat zu kopieren und auf Amazon S3 hochzuladen](#)
- [Schritt 5: Cluster erstellen](#)

## Schritt 1: Sammeln Sie Informationen über Ihren LDAP Server und kopieren Sie das Serverzertifikat auf Amazon S3

Sie benötigen die Informationen und Elemente im folgenden Abschnitt von Ihrem LDAP Server, um die LDAP Authentifizierung zu konfigurieren.

Die IP-Adresse oder der Hostname des LDAP Servers

Der Presto-Koordinator auf dem EMR Amazon-Masterknoten muss in der Lage sein, den LDAP Server unter der angegebenen IP-Adresse oder dem angegebenen Hostnamen zu erreichen. Standardmäßig kommuniziert Presto LDAPS über Port 636 mit dem LDAP Server. Wenn Ihre LDAP Implementierung einen benutzerdefinierten Port erfordert, können Sie ihn mithilfe der `ldap.url` Eigenschaft mit Amazon EMR 5.16.0 oder höher oder `authentication.ldap.url` mit früheren Versionen angeben. Ersetzen Sie 636 durch den benutzerdefinierten Port, wie in den `presto-config`-Konfigurationsklassifizierungsbeispielen unter [Schritt 3: Erstellen Sie eine Konfiguration JSON mit Presto-Eigenschaften für LDAP](#) gezeigt. Stellen Sie sicher, dass alle Firewalls und Sicherheitsgruppen ein- und ausgehenden Datenverkehr über Port 636 (oder den benutzerdefinierten Port) und zudem Port 8446 (oder den benutzerdefinierten Port), der für die interne Clusterkommunikation verwendet wird, zulassen.

Das LDAP Serverzertifikat

Sie müssen die Zertifikatsdatei in einen sicheren Speicherort in Amazon S3 hochladen. Weitere Informationen finden Sie unter [Wie lade ich Dateien und Ordner in einen S3 Bucket hoch?](#) im Benutzerhandbuch für Amazon Simple Storage Service. Sie erstellen eine Bootstrap-Aktion, die dieses Zertifikat aus Amazon S3 auf jeden Knoten im Cluster kopiert, wenn der Cluster gestartet wird. Wählen Sie unter [Schritt 4: Erstellen Sie das Skript, um das LDAP Serverzertifikat zu kopieren und auf Amazon S3 hochzuladen](#). Das Beispielzertifikat ist `s3://MyBucket/ldap_server.crt`.

Die LDAP Servereinstellungen für anonyme Bindungen

Wenn die anonyme Bindung auf PrestoDB deaktiviert ist, benötigen Sie die Benutzer-ID (UID) und das Passwort eines Kontos mit der Berechtigung, sich an den LDAP Server zu binden, damit der PrestoDB-Server eine Verbindung herstellen kann. Sie geben das Kennwort UID und mithilfe der `internal-communication.authentication.ldap.password` Eigenschaften `internal-communication.authentication.ldap.user` und in der Konfigurationsklassifizierung an. `presto-config` Amazon EMR 5.10.0 unterstützt diese Einstellungen nicht, daher muss die anonyme Bindung auf dem LDAP Server unterstützt werden, wenn Sie diese Release-Version verwenden.

Beachten Sie, dass Trino die Konfiguration der anonymen Bindung nicht benötigt.

Um den Status der anonymen Bindung auf dem LDAP Server abzurufen

- Verwenden Sie von einem Linux-Client aus den Befehl [ldapwhoami](#), wie im folgenden Beispiel gezeigt:

```
ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
```

Wenn die anonyme Bindung nicht zulässig ist, gibt der Befehl Folgendes zurück:

```
ldap_bind: Inappropriate authentication (48)
additional info: anonymous bind disallowed
```

Um zu überprüfen, ob ein Konto über Berechtigungen für einen LDAP Server verfügt, der die einfache Authentifizierung verwendet

- Verwenden Sie von einem Linux-Client aus den Befehl [ldapwhoami](#), wie im folgenden Beispiel gezeigt. Das Beispiel verwendet einen fiktiven Benutzer, *presto*, gespeichert auf einem LDAP Open-Server, der auf einer EC2 Instanz mit dem fiktiven Hostnamen läuft *ip-xxx-xxx-xxx-xxx.ec2.internal*. Der Benutzer ist der Organisationseinheit (OU) zugeordnet *admins* und mit dem Passwort *123456*:

```
ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-
xxx-xxx-xxx-xxx.ec2.internal
```

Wenn das Konto gültig ist und über die erforderlichen Berechtigungen verfügt, gibt der Befehl Folgendes zurück:

```
dn:uid=presto,ou=admins,dc=ec2,dc=internal
```

Die Beispielkonfigurationen unter [Schritt 3: Erstellen Sie eine Konfiguration JSON mit Presto-Eigenschaften für LDAP](#) enthalten dieses Konto, um die Verständlichkeit zu verbessern. Ausgenommen ist das 5.10.0-Beispiel, weil es hier nicht unterstützt wird. Wenn der LDAP Server anonyme Bindungen verwendet, entfernen Sie die Paare `internal-communication.authentication.ldap.user` und `internal-communication.authentication.ldap.password` Name/Wert.

## Der definierte LDAP Name (DN) für Presto-Benutzer

Wenn Sie die LDAP Konfiguration für Presto angeben, geben Sie ein Bindungsmuster an, das aus einer `{USER}` Organisationseinheit (OU) und zusätzlichen Domänenkomponenten (DCs) besteht. Presto ersetzt es `{USER}` bei der Kennwortauthentifizierung durch die tatsächliche Benutzer-ID (UID) jedes Benutzers, damit es dem definierten Namen (DN) entspricht, den dieses Bindungsmuster angibt. Sie benötigen OUs die Mitglieder und deren DCs Mitglieder, zu denen berechtigte Benutzer gehören. Um Benutzern aus der `admins`-OU in der `corp.example.com`-Domain zu gestatten, sich bei Presto zu authentifizieren, geben Sie `{USER},ou=admins,dc=corp,dc=example,dc=com` als Benutzerbindungsmuster an.

### Note

Wenn Sie verwenden AWS CloudFormation, müssen Sie die Funktion `Fn::Sub` verwenden, um sie durch die tatsächliche Benutzer-ID (UID) zu `{USER}` ersetzen. Weitere Informationen finden Sie im Thema [Fn::Sub](#) im AWS CloudFormation -Benutzerhandbuch.

Wenn Sie Amazon EMR 5.10.0 verwenden, können Sie nur ein solches Muster angeben. Mit Amazon EMR 5.11.0 oder höher können Sie mehrere Muster angeben, die durch einen Doppelpunkt (:) getrennt sind. Benutzer, die sich bei Presto zu authentifizieren versuchen, werden mit den ersten Muster verglichen, dann mit dem zweiten usw. Ein Beispiel finden Sie unter [Schritt 3: Erstellen Sie eine Konfiguration JSON mit Presto-Eigenschaften für LDAP](#).

## Schritt 2: Sicherheitskonfiguration einrichten

Erstellen Sie eine Sicherheitskonfiguration mit aktivierter Datenverschlüsselung während der Übertragung. Weitere Informationen finden Sie unter [Sicherheitskonfiguration erstellen](#) im Amazon EMR Management Guide. Die Verschlüsselungsartefakte, die Sie bereitstellen, wenn Sie die Verschlüsselung der Daten während der Übertragung einrichten, werden zum Verschlüsseln der internen Kommunikation zwischen Presto-Knoten verwendet. Weitere Informationen finden Sie unter [Bereitstellen von Zertifikaten für die Datenverschlüsselung während der Übertragung](#). Das LDAP Serverzertifikat wird verwendet, um Client-Verbindungen zum Presto-Server zu authentifizieren.

## Schritt 3: Erstellen Sie eine Konfiguration JSON mit Presto-Eigenschaften für LDAP

Sie verwenden die `presto-config` Konfigurationsklassifizierung, um Presto-Eigenschaften für festzulegen. LDAP Das Format und der Inhalt von unterscheiden `presto-config` sich je nach EMR

Amazon-Release-Version und Presto-Installation (PrestoDB oder Trino) geringfügig. Beispiele für Konfigurationsunterschiede finden Sie unten in diesem Abschnitt. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Bei den folgenden Schritten wird davon ausgegangen, dass Sie die JSON Daten in einer Datei speichern, *MyPrestoConfig.json*. Wenn Sie die Konsole verwenden, laden Sie die Datei an einen sicheren Ort in Amazon S3 hoch, damit Sie bei der Erstellung des Clusters darauf verweisen können. Wenn Sie die verwenden AWS CLI, können Sie die Datei lokal referenzieren.

Example Amazon EMR 6.1.0 und höher mit Presto SQL (Trino)

Im folgenden Beispiel wird der LDAP Hostname von verwendet, [Schritt 1: Sammeln Sie Informationen über Ihren LDAP Server und kopieren Sie das Serverzertifikat auf Amazon S3](#) um sich beim Server für die LDAP Bindung zu authentifizieren. Es wurden zwei Benutzerbindungsmuster angegeben, was darauf hinweist, dass Benutzer innerhalb der admins Organisationseinheit und der datascientists Organisationseinheit auf dem LDAP Server für die Authentifizierung beim Trino-Server als Benutzer berechtigt sind. Die Bindungsmuster werden durch einen Doppelpunkt (:) voneinander getrennt.

EMRAmazon-Versionen 6.4.0 und höher verwenden den neuen Namen Trino anstelle von Presto. SQL Wenn Sie Trino verwenden, ersetzen Sie *prestoql-config* in der folgenden Konfigurationsklassifikation mit und mit. trino-config *prestoql-password-authenticator* trino-password-authenticator

```
[
 {
 "Classification": "prestoql-config",
 "Properties": {
 "http-server.authentication.type": "PASSWORD"
 }
 },
 {
 "Classification": "prestoql-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 und höher

Das folgende Beispiel verwendet die LDAP Benutzer-ID und das Passwort sowie den LDAP Hostnamen von, [Schritt 1: Sammeln Sie Informationen über Ihren LDAP Server und kopieren Sie das Serverzertifikat auf Amazon S3](#) um sich beim LDAP Server für die Bindung zu authentifizieren. Es wurden zwei Benutzerbindungsmuster angegeben, die darauf hinweisen, dass Benutzer innerhalb der `admins` Organisationseinheit und der `datascientists` Organisationseinheit auf dem LDAP Server für die Authentifizierung beim Presto-Server als Benutzer berechtigt sind. Die Bindungsmuster werden durch einen Doppelpunkt (:) voneinander getrennt.

```
[{
 "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 bis 5.15.0

Das Format der Konfigurationsklassifizierung `presto-config` unterscheidet sich geringfügig bei diesen Versionen. Das folgende Beispiel gibt dieselben Parameter an, die auch im vorherigen Beispiel verwendet wurden.

```
[{
 "Classification": "presto-config",
 "Properties": {
 "http-server.authentication.type": "LDAP",
 "authentication.ldap.url": "ldaps://ip-xxx-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

Amazon EMR 5.10.0 unterstützt nur anonyme Bindungen, sodass diese Einträge weggelassen werden. Darüber hinaus kann nur ein Bindungsmuster angegeben werden.

```

[{
 "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"
 }
}]

```

## Schritt 4: Erstellen Sie das Skript, um das LDAP Serverzertifikat zu kopieren und auf Amazon S3 hochzuladen

Erstellen Sie ein Skript, das die Zertifikatdatei auf jeden Knoten im Cluster kopiert und in den Schlüsselspeicher einfügt. Erstellen Sie das Skript mit einem Texteditor, speichern Sie es und laden Sie es dann zu Amazon S3 hoch. In [Schritt 5: Cluster erstellen](#) wird auf die Skriptdatei verwiesen als `s3://MyBucket/LoadLDAPCert.sh`.

Das folgende Beispielskript verwendet das Standard-Keystore-Passwort: `changeit`. Wir empfehlen, dass Sie nach der Erstellung des Clusters eine Verbindung zum Master-Knoten herstellen und das Keystore-Passwort mit dem Befehl `keytool` ändern.

```

#!/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

```

## Schritt 5: Cluster erstellen

Wenn Sie den Cluster erstellen, geben Sie Presto und andere Anwendungen an, die Amazon installieren EMR soll. Die folgenden Beispiele verweisen auch auf die Eigenschaften der Konfigurationsklassifizierung in aJSON, aber Sie können die Konfigurationsklassifizierung auch direkt angeben.

Um einen Presto-Cluster mit LDAP Authentifizierung mithilfe der EMR Amazon-Konsole zu erstellen

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie Presto zusammen mit anderen Anwendungen, die Amazon installieren EMR soll, und wählen Sie unter Softwarekonfiguration die Version von Amazon aus, die Sie verwenden EMR möchten. LDAP Die Authentifizierung wird nur mit Amazon EMR 5.10.0 und höher unterstützt.
4. Wählen Sie unter Softwareeinstellungen bearbeiten die Option JSON Aus S3 laden, geben Sie den Speicherort der JSON Konfigurationsdatei, in der Sie sie erstellt haben [Schritt 3: Erstellen Sie eine Konfiguration JSON mit Presto-Eigenschaften für LDAP](#), in Amazon S3 ein und wählen Sie dann Weiter.
5. Konfigurieren Sie die Cluster-Hardware und die Netzwerkfunktionen. Wählen Sie dann Next (Weiter).
6. Wählen Sie Bootstrap Actions (Bootstrap-Aktionen). Wählen Sie für Add bootstrap action (Bootstrap-Aktion hinzufügen) die Option Custom action (Benutzerdefinierte Aktion) und wählen Sie dann Configure and add (Konfigurieren und hinzufügen).
7. Geben Sie einen Namen für die Bootstrap-Aktion ein, geben Sie den Skriptspeicherort ein, in dem Sie das Skript erstellt haben [Schritt 4: Erstellen Sie das Skript, um das LDAP Serverzertifikat zu kopieren und auf Amazon S3 hochzuladen](#), z. B. s3://MyBucket/LoadLDAPCert.sh, und wählen Sie dann Hinzufügen.
8. Wählen Sie unter General Options (Allgemeine Optionen), Tags und Additional Options (Weitere Optionen) für die Anwendung geeignete Einstellungen. Klicken Sie dann auf Next (Weiter).
9. Wählen Sie Authentication and encryption (Authentifizierung und Verschlüsselung) und dann die Security configuration (Sicherheitskonfiguration), die Sie in [Schritt 2: Sicherheitskonfiguration einrichten](#) erstellt haben.

- Wählen Sie weitere für Ihre Anwendung geeignete Sicherheitsoptionen und wählen Sie dann `Create cluster` (Cluster erstellen).

Um einen Presto-Cluster mit LDAP Authentifizierung zu erstellen, verwenden Sie den AWS CLI

- Verwenden Sie den `aws emr create-cluster`-Befehl. Geben Sie mindestens die Presto-Anwendung und auch die Presto-Konfigurationsklassifizierung, das Bootstrap-Skript und die Sicherheitskonfiguration an, die Sie in den vorherigen Schritten erstellt haben. Das folgende Beispiel verweist auf die Konfigurationsdatei als JSON Datei, die in demselben Verzeichnis gespeichert ist, in dem Sie den Befehl ausführen. Das Bootstrap-Skript muss dagegen in Amazon S3 gespeichert werden. Im folgenden Beispiel wird verwendet `s3://MyBucket/LoadLDAPCert.sh`.

#### Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

## Den strikten Presto-Modus aktivieren

In bestimmten Situationen können lang andauernde Abfragen zu hohen Kosten führen und Amazon veranlassen, mehr Cluster-Ressourcen EMR zu verwenden. Dadurch werden anderen Workloads im Cluster Ressourcen entzogen. Mit EMR Amazon-Versionen 6.8 und höher können Sie eine Funktion im strikten Modus verwenden, die die folgenden Arten von Abfragen mit langer Laufzeit ablehnt oder Sie davor warnt:

- Abfragen ohne Prädikate für die partitionierten Spalten, die zu Tabellenscans großer Datenmengen führen
- Abfragen mit Kreuzverknüpfungen zwischen zwei großen Tabellen
- Abfragen, die eine große Anzahl von Zeilen ohne Limit sortieren

Nachdem Presto den Abfrageplan vollständig optimiert hat, wird der strikte Modus ausgeführt. Um den strikten Modus zu verwenden und an Ihre Abfrageanforderungen anzupassen, können Sie Presto wie folgt konfigurieren.

### Presto-Konfigurationen für den strikten Modus

Einstellung	Beschreibung	Standard
<code>strict-mode-enabled</code>	Schaltet den strikten Modus ein und aus. Der Wert von <code>true</code> gibt an, dass der strikte Modus aktiviert ist.	<code>false</code>
<code>strict-mode-fail-query</code>	Weist Abfragen ab, wenn der strikte Modus Abfragen mit wahrscheinlich langer Laufzeit erkennt. Wenn <code>false</code> , gibt Amazon EMR nur eine Warnung aus.	<code>false</code>
<code>strict-mode-restrictions</code>	Gibt die Einschränkungen an, die gelten sollen, wenn der strikte Modus aktiviert ist. Der strikte Modus unterstützt die folgenden Einschränkungen: <code>MANDATORY_PARTITION_DISALLOW_PREDICATEJOIN</code> , <code>CROSS_JOIN</code> und <code>LIMITED_SORT</code> .	<code>MANDATORY_PARTITION_PREDICATE, DISALLOW_CROSS_JOIN, LIMITED_SORT</code>

Um mit dem strikten Modus zu experimentieren, können Sie diese Konfigurationen überschreiben oder sie als Sitzungseigenschaften festlegen, wenn Sie den Presto-Client verwenden.

Um die Konfiguration bei der Clustererstellung festzulegen, verwenden Sie den AWS Management Console

1. Wählen Sie Cluster erstellen und wählen Sie Amazon EMR Version 6.8.0 und Presto oder Trino aus. Weitere Informationen finden Sie unter [Installieren von PrestoDB und Trino](#).
2. Geben Sie die Konfigurationseigenschaften für den strikten Modus direkt an oder laden Sie eine JSON Datei auf Amazon S3 hoch. Wählen Sie optional den für Ihren Metastore aus. Geben Sie Ihre SubnetzeVPC, Bootstrap-Aktionen, Ihr key pair und Ihre Sicherheitsgruppe an. Um Ihren Cluster jetzt zu erstellen, wählen Sie Cluster erstellen aus.
3. Melden Sie sich beim Primärknoten des Clusters an und führen Sie `presto-cli` oder `trino-cli` aus.
4. Reichen Sie Ihre Anfragen ein. Der strikte Modus validiert jede Abfrage und stellt fest, ob sie lange andauert. Abhängig von Ihrer `strict-mode-fail-query` Einstellung EMR lehnt Amazon die Anfrage ab oder gibt eine Warnung aus.
5. Wenn Sie mit Ihren Abfragen fertig sind, beenden Sie den Cluster und löschen Sie Ihre Ressourcen.

Um die Konfiguration auf einem laufenden Cluster festzulegen, verwenden Sie AWS CLI

1. Melden Sie sich mit dem am primären Knoten Ihres Clusters an AWS CLI und führen Sie `presto-cli` oder `austrino-cli`.
2. Führen Sie die folgenden Befehle mit den gewünschten Werten aus.

```
set session strict_mode_enabled = true;
set session strict_mode_fail_query = false;
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

## Überlegungen

Berücksichtigen Sie bei der Verwendung des strikten Modus die folgenden Punkte:

- In einigen Fällen kann der strikte Modus Abfragen mit kurzer Laufzeit zurückweisen, die nicht viele Ressourcen verbrauchen. Beispielsweise wenden Abfragen in kleinen Tabellen keine dynamische Filterung an und ersetzen auch keine inneren Verknüpfungen durch Kreuzverknüpfungen. Dies kann dazu führen, dass die Abfrage das obligatorische Partitionsprädikat verwendet oder Cross-Joins nicht zulässt. In diesem Fall lehnt der strikte Modus die Abfrage ab.

- Die strenge Modusprüfung wird nur für die Typen SELECT INSERTSELECT, CREATE TABLE AS und EXPLAIN ANALYZE Abfragetypen angewendet.
- Sie können die Einschränkung für das obligatorische Partitionsprädikat nur mit dem Hive-Konnektor verwenden.

## Umgang mit dem Verlust von Spot Instances in Presto

Mit Spot-Instances in Amazon EMR können Sie Big-Data-Workloads auf freien EC2 Amazon-Kapazitäten zu reduzierten Kosten ausführen. Im Gegenzug für die geringeren Kosten EC2 kann Amazon Spot-Instances mit einer zweiminütigen Benachrichtigung unterbrechen. Wenn Sie einen Knoten beenden, kann es bis zu 10 Minuten dauern, bis Presto einen Fehler zurückgibt. Dies führt zu unnötigen Verzögerungen bei Fehlerberichten und möglichen Wiederholungsversuchen. Schnelle Beendigung ist ein Feature, mit der Sie kontrollieren können, wie Presto mit beendeten Knoten umgeht.

Die Aufgabe des Presto-Koordinators besteht darin, den Überblick über alle Worker-Knoten zu behalten und deren Status regelmäßig abzufragen. Ohne eine schnelle Kündigung konsultiert der Koordinator den YARN NodeManager Status der einzelnen Knoten nicht. Dies kann zu einer langen Wiederholungsschleife führen, bevor die Abfrage fehlschlägt. Bei einer Schnellbeendigung fragt der Presto-Koordinator den Knotenstatus ab, NodeManager sobald die Abfrage den Host nicht erreicht. Wenn sich NodeManager herausstellt, dass der Knoten inaktiv ist, bricht Presto weitere Versuche ab, schlägt die Abfrage fehl und gibt einen Fehler zurück. `NODE_DECOMMISSIONED`

Mit den folgenden Konfigurationsparametern können Sie das Verhalten von Presto im Falle einer Knotenbeendigung steuern und anpassen.

### Presto-Konfigurationen für die Behandlung von Knotenausfällen

Einstellung	Beschreibung	Standard
<code>query.remote-task.max-backoff-duration</code>	Die Dauer, während der der Koordinator weiterhin versucht, den Status der Remote-Aufgabe von den Worker-Knoten abzurufen.	10 Minuten

Einstellung	Beschreibung	Standard
<code>query.remote-task.quick-terminate-no-de-failure</code>	<p>Aktiviert den schnellen Knotenausfall, wenn der Koordinator den Knoten nicht erreichen oder keine Verbindung zu dem Worker herstellen kann, der auf diesem Knoten ausgeführt wird. Der Wert von <code>query.remote-task.terminate-on-connection-exception</code> bestimmt, ob der Koordinator den Knoten erreichen oder eine Verbindung zum Worker herstellen muss.</p> <p>Der Knoten schlägt die Abfrage fehl und Amazon EMR entfernt den Knoten aus der Liste der verfügbaren Worker. In diesem Fall können Sie den Knoten nicht verwenden, um neue Abfragen zu planen.</p> <p>Wenn Sie diesen Wert auf <code>false</code> setzen, kehrt Presto zu seinem vorherigen Verhalten zurück, bei dem der Presto-Koordinator erneut versucht, den Knoten (für <code>query.remote-task.max-backoff-duration</code>) zu erreichen, bevor er den Knoten als nicht verfügbar markiert und die</p>	<code>true</code>

Einstellung	Beschreibung	Standard
	laufende Abfrage auf dem Knoten fehlschlägt.	
<code>query.remote-task.terminate-on-connection-exception</code>	Gibt an, ob Amazon einen Knoten einrichten EMR soll, wenn der Host erreichbar ist, der Koordinator aber keine Verbindung zum Worker-Pr o z e s s des Hosts herstellen kann. Wenn Sie diesen Wert auf <code>true</code> setzen, aktivieren Sie „schneller Abfragefehler“, falls der Host nicht erreichbar ist.	<code>false</code>

## Fehlertolerante Ausführung in Trino

Die fehlertolerante Ausführung ist ein Mechanismus in Trino, den ein Cluster verwenden kann, um Abfragefehler zu minimieren. Zu diesem Zweck werden Abfragen oder deren Komponentenaufgaben wiederholt, wenn sie fehlschlagen. Wenn die fehlertolerante Ausführung aktiviert ist, werden die zwischengeschalteten Austauschdaten gespooled, sodass sie im Falle eines Worker-Ausfalls oder eines anderen Fehlers während der Abfrageausführung von einem anderen Worker wiederverwendet werden können.

Weitere Informationen zur fehlertoleranten Ausführung in Trino finden Sie im Trino-Blog unter [Project Tardigrade delivers ETL at Trino-speed to early users](#).

## Konfiguration

Die fehlertolerante Ausführung ist standardmäßig deaktiviert. Um das Feature zu aktivieren, legen Sie die `retry-policy` Konfigurationseigenschaft in der `trino-config`-Klassifizierung wie folgt auf QUERY oder TASK basierend auf der gewünschten Wiederholungsrichtlinie fest.

```
{"classification":
 "trino-config",
"properties":
```



```
{
 "retry-policy":
 "QUERY"
}
```

Eine **QUERY**Wiederholungsrichtlinie weist Trino an, eine Abfrage automatisch zu wiederholen, wenn auf einem Worker-Knoten ein Fehler auftritt. Wir empfehlen, eine QUERY Wiederholungsrichtlinie zu verwenden, wenn der Großteil der Workload für den Trino-Cluster aus vielen kleinen Abfragen besteht.

Eine **TASK**Wiederholungsrichtlinie weist Trino an, einzelne Abfrageaufgaben im Falle eines Fehlers erneut zu versuchen. Wir empfehlen diese Richtlinie, wenn Trino umfangreiche Batch-Abfragen ausführt. Der Cluster kann kleinere Aufgaben innerhalb der Abfrage effizienter wiederholen, als die gesamte Abfrage erneut zu versuchen.

## Austausch-Manager

Ein Exchange-Manager speichert und verwaltet Spooldaten für eine fehlertolerante Ausführung. Er verwendet externen Speicher, um verschüttete Daten zu speichern, die über die Größe des In-Memory-Puffers hinausgehen. Sie können einen dateisystembasierten Exchange-Manager konfigurieren, der Spooldaten an einem bestimmten Ort speichert, z. B. Amazon S3, Amazon S3 S3-kompatible Systeme oder. HDFS

Die EMR Amazon-Versionen 6.9.0 und höher enthalten die `trino-exchange-manager` Klassifizierung zur Konfiguration des Exchange Managers. Diese Versionen unterstützen auch HDFS Spooling.

### Exchange Manager einrichten

Verwenden Sie die `trino-exchange-manager`-Konfigurationsklassifizierung, um einen Exchange Manager zu konfigurieren. Diese Klassifizierung erstellt intern eine `etc/exchange-manager.properties`-Konfigurationsdatei für den Koordinator und alle Worker-Knoten. Verwenden Sie die Konfigurationsklassifizierung , um die `exchange-manager.name`-Eigenschaften auf `filesystem` festzulegen.

Standardmäßig werden EMR Amazon-Versionen 6.9.0 und höher HDFS als Exchange-Manager verwendet. HDFS ist in den EMR EC2 Amazon-Clustern verfügbar, und das Spooling erfolgt standardmäßig im `trino-exchange/` Verzeichnis. Um die Standardeinstellungen zu verwenden, legen Sie die folgende Konfiguration fest:

```

{"Classification":
 "trino-exchange-manager"
}

```

Wenn Sie einen benutzerdefinierten Speicherort angeben möchten, legen Sie die folgenden Eigenschaften in der `trino-exchange-manager`-Klassifizierung fest:

- Setzen Sie `exchange.use-local-hdfs` auf `true`.
- Legen Sie `exchange.base-directories` den Speicherort für das benutzerdefinierte Verzeichnis festHDFS, z. B. in `exchange.base-directories=/exchange` Wenn das benutzerdefinierte Verzeichnis noch nicht vorhanden istHDFS, erstellt Amazon EMR es.

### HDFSExchange Manager-Konfigurationen

Auf der Grundlage interner Testergebnisse empfehlen wir, HDFS für eine bessere Abfrageleistung im Vergleich zu anderen cloudbasierten Dateisystemen den Spoolvorgang auf lokal zu setzen. Mit können Sie die folgenden Konfigurationen für den Exchange-Manager festlegenHDFS.

Konfiguration	Beschreibung	Standardeinstellung
<code>exchange.hdfs.block-size</code>	Blockgröße für die HDFS Speicherung	4 MB
<code>hdfs.config.resources</code>	Liste der zu konfigurierenden Dateipfade HDFS	Falls <code>exchange.use-local-hdfs true</code> ist, werden die Pfade zu <code>core-site.xml</code> , <code>hdfs-site.xml</code> Dateien verwendet, andernfalls <code>null</code>

Zusätzliche Konfigurationseigenschaften für die fehlertolerante Ausführung und Informationen zur Einrichtung von Amazon S3 oder anderen Amazon-S3-kompatiblen Systemen für das Spooling finden Sie auf der Seite [Fehlertolerante Ausführung](#) in der Trino-Dokumentation.

## Überlegungen und Einschränkungen

- Wenn Sie die fehlertolerante Ausführung aktivieren, werden `write`-Operationen für Konnektoren deaktiviert, die `write` nicht unterstützen, wenn `retry-policy` festgelegt ist. Ab EMR Amazon-Version 6.9.0 unterstützen `write` Delta Lake-, Hive- und Iceberg-Konnektoren den Betrieb mit `retry-policy`.
- Wenn Sie Exchange Manager verwenden und teure E/A-Vorgänge durchführen, kann es bei Ihren Abfragen zu Leistungseinbußen kommen, während Exchange Manager die Zwischendaten auf externen Speicher überträgt.

## Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden

Die EMR Versionen 5.30.0 und höher von Amazon enthalten eine Funktion, mit der Sie eine Übergangsfrist für bestimmte Skalierungsaktionen festlegen können. Die Übergangszeit ermöglicht es, dass Presto-Aufgaben weiterlaufen, bevor der Knoten aufgrund einer Skalierungsaktion zur Größenänderung oder einer automatischen Skalierungsrichtlinienanforderung beendet wird. Weitere Informationen zu Skalierungsregeln finden Sie unter [Grundlegendes zu automatischen Skalierungsregeln](#) im Amazon EMR Management Guide. Auto Scaling in Presto mit ordnungsgemäßer Stilllegung verhindert, dass neue Aufgaben auf einem Knoten geplant werden, der außer Betrieb genommen wird, und ermöglicht gleichzeitig, dass bereits laufende Aufgaben abgeschlossen werden, bevor die Zeitüberschreitung für das Herunterfahren erreicht ist. Derzeit laufende Abfragen schließen die Ausführung ab, bevor der Knoten stillgelegt wird. Auto Scaling wird für Instance-Flotten nicht unterstützt.

Sie können steuern, wie viel Zeit für die Ausführung von Presto-Aufgaben eingeräumt werden soll, nachdem eine Anfrage zum Herunterfahren der automatischen Skalierung eingegangen ist. Standardmäßig EMR beträgt das Shutdown-Timeout für Amazon 0 Minuten, was bedeutet, dass Amazon den Knoten und alle darauf ausgeführten Presto-Aufgaben EMR sofort beendet, falls dies für eine Scale-In-Anfrage erforderlich ist. Um ein längeres Timeout für Presto-Aufgaben auf Amazon festzulegen EMR, damit laufende Abfragen abgeschlossen werden können, bevor ein Cluster herunterskaliert wird, verwenden Sie die `presto-config` Konfigurationsklassifizierung, um den `graceful-shutdown-timeout` Parameter auf einen Wert in Sekunden oder Minuten über Null festzulegen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Wenn Sie beispielsweise den Wert `graceful-shutdown-timeout` auf `"30m"` erhöhen, wird ein Zeitraum von 30 Minuten angegeben. Nach Ablauf der Zeitüberschreitung für das Herunterfahren wird der Knoten, der für die Stilllegung markiert ist, zwangsweise beendet, wenn er auf den Abschluss von Abfrageaufgaben wartet und die Abfrage fehlschlägt. Wenn die Abfrageaufgaben in fünf Minuten abgeschlossen sind, wird der für die Außerbetriebnahme markierte Knoten nach fünf Minuten beendet, sofern die Ausführung anderer YARN Anwendungen abgeschlossen ist.

Example Beispiel einer automatischen Skalierung in Presto mit ordnungsgemäßer Stilllegung

Ersetzen Sie den `graceful-shutdown-timeout`-Wert durch die Anzahl der Minuten, die für Ihre Einrichtung geeignet ist. Es gibt keinen Höchstwert. Im folgenden Beispiel wird ein Timeout-Wert von 1800 Sekunden (30 Minuten) festgelegt.

```
[
 {
 "classification": "presto-config",
 "properties": {
 "graceful-shutdown-timeout": "1800s"
 }
 }
]
```

## Einschränkungen

PrestoDB Graceful Decommission funktioniert nicht auf EMR Clustern, bei denen die HTTP Konnektivität deaktiviert ist, z. B. wenn sie `http-server.http.enabled` auf `false` eingestellt ist. Trino unterstützt Ordnungsgemäße Stilllegung, unabhängig von der Einstellung `http-server.http.enabled`, überhaupt nicht.

## Überlegungen zu Presto bei Amazon EMR

Beachten Sie die folgenden Einschränkungen, wenn Sie [Presto](#) auf Amazon EMR ausführen.

### In der Befehlszeile von Presto ausführbare Datei

In Amazon EMR verwenden PrestoDB und Trino beide dieselbe ausführbare Befehlszeilendatei `presto-cli`, wie im folgenden Beispiel.

```
presto-cli --catalog hive
```

## Nicht konfigurierbare Presto-Bereitstellungseigenschaften

Die Version von Amazon EMR, die Sie verwenden, bestimmt die verfügbaren Presto-Bereitstellungskonfigurationen. Weitere Informationen zu diesen Konfigurationseigenschaften finden Sie unter [Bereitstellen von Presto](#) in der Presto-Dokumentation. Die folgende Tabelle zeigt die Konfigurationsoptionen der Presto-properties-Dateien.

Datei	Konfigurierbar
log.properties	<p>PrestoDB: Konfigurierbar in EMR Amazon-Versionen 4.0.0 und höher. Verwenden Sie die <code>presto-log</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in EMR Amazon-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-log</code> - oder <code>trino-log</code> -Konfigurationsklassifizierung.</p>
config.properties	<p>PrestoDB: Konfigurierbar in EMR Amazon-Versionen 4.0.0 und höher. Verwenden Sie die <code>presto-config</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in EMR Amazon-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-config</code> - oder <code>trino-config</code> -Konfigurationsklassifizierung.</p>
hive.properties	<p>PrestoDB: Konfigurierbar in EMR Amazon-Versionen 4.1.0 und höher. Verwenden Sie die <code>presto-connector-hive</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in EMR Amazon-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-connector</code></p>

Datei	Konfigurierbar
	r-hive - oder trino-connector-hive - Konfigurationsklassifizierung.
node.properties	PrestoDB: Konfigurierbar in EMR Amazon-Version 5.6.0 und höher. Verwenden Sie die presto-node -Konfigurationsklassifizierung.  Trino (PrestoSQL): Konfigurierbar in EMR Amazon-Versionen 6.1.0 und höher. Verwenden Sie die prestosql-node - oder trino-node -Konfigurationsklassifizierung.
jvm.config	Nicht konfigurierbar.

## Installation von PrestoDB und Trino

Der Anwendungsname Presto wird weiterhin verwendet, um PrestoDB auf Clustern zu installieren. Um Trino auf Clustern zu installieren, verwenden Sie den Anwendungsnamen Trino (oder Presto SQL in älteren Versionen von Amazon). EMR

Sie können entweder PrestoDB oder Trino installieren, aber Sie können nicht beide auf einem einzigen Cluster installieren. Wenn Sie beim Versuch, einen Cluster zu erstellen, sowohl PrestoDB als auch Trino angeben, tritt ein Validierungsfehler auf und die Anfrage zur Clustererstellung schlägt fehl.

## EMRFS und PrestoS3-Konfiguration File System

Mit EMR Amazon-Versionen 5.12.0 und höher kann PrestoDB verwendet werden. EMRFS Dies ist die Standardkonfiguration. EMRFS ist auch das Standarddateisystem für Trino (PrestoSQL) in EMR Amazon-Versionen 6.1.0 und höher. Weitere Informationen finden Sie unter [EMRDateisystem \(EMRFS\)](#) im Amazon EMR Management Guide. In früheren Versionen von Amazon EMR File System ist PrestoS3 die einzige Konfigurationsoption.

Sie können eine Sicherheitskonfiguration verwenden, um die Verschlüsselung von EMRFS Daten in Amazon S3 einzurichten. Sie können IAM Rollen auch für EMRFS Anfragen an Amazon S3 verwenden. Weitere Informationen finden Sie unter [Grundlegendes zu Verschlüsselungsoptionen](#) und

## [Konfigurieren von IAM Rollen für EMRFS Anfragen an Amazon S3](#) im Amazon EMR Management Guide.

### Note

Wenn Sie die zugrunde liegenden Daten in Amazon S3 mit Amazon EMR Version 5.12.0 abfragen, können Presto-Fehler auftreten. Dies liegt daran, dass Presto die Konfigurationsklassifizierungswerte von `emrfs-site.xml` nicht abrufen kann. Um dieses Problem zu umgehen, erstellen Sie ein `emrfs`-Unterverzeichnis unter `usr/lib/presto/plugin/hive-hadoop2/` und erstellen einen Symlink `usr/lib/presto/plugin/hive-hadoop2/emrfs` zu der vorhandenen Datei `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml`. Starten Sie dann den Presto-Server-Prozess neu (`sudo presto-server stop` gefolgt von `sudo presto-server start`).

Sie können die EMRFS Standardeinstellung überschreiben und stattdessen PrestoS3 verwenden. Filesystem Verwenden Sie dazu die `presto-connector-hive`-Konfigurationsklassifizierung, um `hive.s3-file-system-type` auf `PRESTO` zu setzen, wie im folgenden Beispiel veranschaulicht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
[
 {
 "Classification": "presto-connector-hive",
 "Properties": {
 "hive.s3-file-system-type": "PRESTO"
 }
 }
]
```

Wenn Sie PrestoS3 verwenden, verwenden Sie die `presto-connector-hive` Konfigurationsklassifizierung oder `trino-connector-hive` für TrinoFilesystem, um PrestoS3-Eigenschaften zu konfigurieren. Filesystem Weitere Informationen zu den verfügbaren Eigenschaften finden Sie unter [Amazon-S3-Konfiguration](#) im Abschnitt zum Hive-Konnektor in der Presto-Dokumentation. Diese Einstellungen gelten nicht für EMRFS.

## Standardeinstellung für Endbenutzer-Identitätswechsel

Standardmäßig ermöglichen EMR Amazon-Versionen 5.12.0 und höher Endbenutzer-Identitätswechsel für den Zugriff auf HDFS. Weitere Informationen finden Sie im

Abschnitt zum [Endbenutzer-Identitätswechsel](#) in der Presto-Dokumentation. Um diese Einstellung zu ändern, setzen Sie die `presto-config`-Konfigurationsklassifizierung der `hive.hdfs.impersonation.enabled`-Eigenschaft auf `false`.

## Standardport für die Presto-Webschnittstelle

Standardmäßig EMR konfiguriert Amazon die Presto-Weboberfläche auf dem Presto-Koordinator für die Verwendung von Port 8889 (für PrestoDB und Trino). Um den Port zu ändern, verwenden Sie die `presto-config`-Konfigurationsklassifizierung, um die `http-server.http.port`-Eigenschaft festzulegen. Weitere Informationen finden Sie unter [Konfigurationseigenschaften](#) im Abschnitt zur Bereitstellung von Presto in der Presto-Dokumentation.

## Problem bei der Hive-Bucket-Ausführung in einigen Versionen

Presto Version 152.3 hat ein Problem mit der Hive-Bucket-Ausführung, das in Presto in einigen Fällen zu einer erheblich langsameren Abfrageleistung führt. Die EMR Amazon-Versionen 5.0.3, 5.1.0 und 5.2.0 enthalten diese Version von Presto. Um dieses Problem abzumildern, verwenden Sie die `presto-connector-hive`-Konfigurationsklassifizierung, um die Eigenschaft `hive.bucket-execution` auf `false` zu setzen, wie im folgenden Beispiel veranschaulicht.

```
[
 {
 "Classification": "presto-connector-hive",
 "Properties": {
 "hive.bucket-execution": "false"
 }
 }
]
```

## Presto-Versionsverlauf

In der folgenden Tabelle sind die in jeder Release-Version von Amazon EMR enthaltenen Presto-Versionen sowie die mit der Anwendung installierten Komponenten aufgeführt. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).



## Die Presto-Versionsinformationen

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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-

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
		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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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



Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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



Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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



Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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
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

Amazon EMR Release-Etikett	Presto-Version	Mit Presto installierte Komponenten
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

In der folgenden Tabelle sind die in jeder Release-Version von Amazon EMR enthaltenen Versionen von Trino (PrestoSQL) sowie die mit der Anwendung installierten Komponenten aufgeführt. Presto SQL änderte seinen Namen ab Version 351 in Trino.

#### Versionsinformationen zu Trino (PrestoSQL)

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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,

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.0.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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



Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
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-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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



Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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
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

Amazon EMR Release-Etikett	Trino (PrestoSQL) Ausführung	Mit Trino (Presto) installierte Komponenten SQL
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

## Versionshinweise zu Trino (PrestoSQL) nach Version

- [Amazon EMR 6.9.0 — Versionshinweise zu Trino \(PrestoSQL\)](#)

### Amazon EMR 6.9.0 — Versionshinweise zu Trino (PrestoSQL)

#### Amazon EMR 6.9.0 — Neue Funktionen von Trino (PrestoSQL)

- Um Abfragen mit langer Laufzeit zu unterstützen, verfügt Trino jetzt über einen fehlertoleranten Ausführungsmechanismus. Die fehlertolerante Ausführung minimiert Abfragefehler, indem fehlgeschlagene Abfragen oder deren Komponentenaufgaben wiederholt werden. Weitere Informationen finden Sie unter [Fehlertolerante Ausführung in Trino](#).

#### Amazon EMR 6.9.0 — Änderungen an Trino (SQLPresto)

#### Amazon EMR 6.9.0 — PrestoDB-Änderungen

Typ	Beschreibung
Upgrade	Upgrade auf PrestoDB 0.276

Typ	Beschreibung
Upgrade	Unterstützung von Hadoop 3.3.3
Upgrade	Hudi-Upgrade auf 0.12.1
Funktion	Amazon EMR - und Presto-Integration mit AWS Lake Formation für interaktive Workloads verwenden. GCSC API
Funktion	Kerberos-bezogene Konfigurationen wurden in Sicherheitskonfigurationen für PrestoDB hinzugefügt, um Kerberos zu aktivieren.
Fehlerbehebung	Die OSS Pull-Anfrage #18115, die hinzugefügt wurde, um die Anzahl der Kopien zu reduzieren, wurde rückgängig gemacht. <code>hdfsConfiguration</code> Dies führte dazu, dass die HDFS Konfiguration bei der Verwendung von EMRFS Hudi-Tabellen falsch kopiert wurde.

### Amazon EMR 6.9.0 — Trino ändert sich

Typ	Beschreibung
Upgrade	Trino-Upgrade auf 398
Upgrade	Unterstützung von Hadoop 3.3.3
Funktion	Tardigrade-Unterstützung: Unterstützung für Exchange-Spooling auf HDFS Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter <a href="#">Fehlertolerante Ausführung in Trino</a>
Fehlerbehebung	Wenn Trino Iceberg verwendet wird und der Glue-Katalog aktiviert ist, vermeiden Sie das Hinzufügen von Metastore-URI in <code>iceberg.properties</code> .

## Amazon EMR 6.9.0 — Bekannte Probleme mit Trino (PrestoSQL)

- Für die EMR Amazon-Version 6.9.0 funktioniert Trino nicht auf Clustern, die für Apache Ranger aktiviert sind. Wenn Sie Trino mit Ranger verwenden müssen, wenden Sie sich an [AWS Support](#).

# Apache Spark

[Apache Spark](#) ist ein Framework und Programmiermodell für verteilte Verarbeitung, mit dem Sie maschinelles Lernen, Stream-Verarbeitung oder Diagrammanalysen mit EMR Amazon-Clustern durchführen können. Ähnlich wie Apache Hadoop ist Spark ein verteiltes Open-Source-Verarbeitungssystem, das häufig für Big-Data-Workloads verwendet wird. Spark weist jedoch einige bemerkenswerte Unterschiede zu Hadoop MapReduce auf. Spark verfügt über eine optimierte Engine zur Ausführung gerichteter azyklischer Graphen (DAG) und speichert Daten aktiv im Speicher, wodurch die Leistung gesteigert werden kann, insbesondere bei bestimmten Algorithmen und interaktiven Abfragen.

Spark unterstützt standardmäßig Anwendungen, die in Scala, Java und Python geschrieben sind. Es enthält auch mehrere eng integrierte Bibliotheken für SQL ([Spark SQL](#)), maschinelles Lernen ([MLlib](#)), Stream-Verarbeitung ([Spark-Streaming](#)) und Graphverarbeitung ([GraphX](#)). Diese Tools vereinfachen die Nutzung des Spark-Frameworks für eine Vielzahl von Anwendungsfällen.

Sie können Spark zusammen mit anderen Hadoop-Anwendungen auf einem EMR Amazon-Cluster installieren und es kann auch das EMR Amazon-Dateisystem (EMRFS) nutzen, um direkt auf Daten in Amazon S3 zuzugreifen. Hive ist auch in Spark integriert, sodass Sie ein HiveContext Objekt verwenden können, um Hive-Skripte mit Spark auszuführen. Ein Hive-Kontext ist als `sqlContext` Bestandteil der Spark-Shell.

Ein Beispiel-Tutorial zur Einrichtung eines EMR Clusters mit Spark und zur Analyse eines Beispieldatensatzes finden Sie unter [Tutorial: Erste Schritte mit EMR Amazon](#) im AWS News-Blog.

## Important

[Apache Spark Version 2.3.1, verfügbar ab EMR Amazon-Version 5.16.0, adressiert CVE-2018-8024 und -2018-1334. CVE](#) Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

In der folgenden Tabelle sind die Version von Spark aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Spark EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Spark installiert wurden, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

## Spark-Versionsinformationen für emr-7.2.0

EMRAmazon-Freigabeetikett	Spark-Version	Mit Spark installierte Komponenten
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


In der folgenden Tabelle sind die Version von Spark aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon zusammen mit Spark EMR installiert.

Die Version der Komponenten, die mit Spark in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

## Spark-Versionsinformationen für emr-6.15.0

EMRAmazon-Freigabeetikett	Spark-Version	Mit Spark installierte Komponenten
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,

EMR Amazon-Freigabeetikett	Spark-Version	Mit Spark installierte Komponenten
		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

 Note

EMR Amazon-Version 6.8.0 wird mit Apache Spark 3.3.0 geliefert. Diese Spark-Version verwendet Apache Log4j 2 und die `log4j2.properties`-Datei zur Konfiguration von Log4j in Spark-Prozessen. Wenn Sie Spark im Cluster verwenden oder EMR Cluster mit benutzerdefinierten Konfigurationsparametern erstellen und ein Upgrade auf EMR Amazon-Version 6.8.0 durchführen möchten, müssen Sie auf die neue `spark-log4j2` Konfigurationsklassifizierung und das neue Schlüsselformat für Apache Log4j 2 migrieren. Weitere Informationen finden Sie unter [Migration von Apache Log4j 1.x zu Log4j 2.x](#).

In der folgenden Tabelle sind die Version von Spark aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Spark EMR installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Spark installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

Spark-Versionsinformationen für emr-5.36.2

EMR Amazon-Freigabeetikett	Spark-Version	Mit Spark installierte Komponenten
emr-5.36.2	Spark 2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-

EMRAmazon-Freigabeetikett	Spark-Version	Mit Spark installierte Komponenten
		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

## Themen

- [Einen Cluster mit Apache Spark erstellen](#)
- [Führen Sie Spark-Anwendungen mit Docker auf Amazon EMR 6.x aus](#)
- [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#)
- [Konfigurieren von Spark](#)
- [Die Leistung von Spark optimieren](#)
- [Caching von Ergebnisfragmenten in Spark](#)
- [Verwenden Sie den Nvidia RAPIDS Accelerator für Apache Spark](#)
- [Zugriff auf die Spark-Shell](#)
- [Verwenden Sie Amazon SageMaker Spark für maschinelles Lernen](#)
- [Eine Spark-Anwendung schreiben](#)
- [Verbessern der Spark Leistung mit Amazon S3](#)
- [Einen Spark-Schritt hinzufügen](#)
- [Anzeigen des Spark-Anwendungsverlaufs](#)
- [Greifen Sie auf das Spark-Web zu UIs](#)



- [Verwenden des Amazon Kinesis Data Streams-Connectors für strukturiertes Streaming mit Spark](#)
- [Verwenden der Amazon Redshift Redshift-Integration für Apache Spark mit Amazon EMR](#)
- [Spark-Versionsverlauf](#)

## Einen Cluster mit Apache Spark erstellen

Mit dem folgenden Verfahren wird mithilfe von Quick Options in der EMR Amazon-Konsole ein Cluster erstellt, auf dem [Spark](#) installiert ist.

Sie können alternativ Erweiterte Optionen verwenden, um Ihre Cluster-Einrichtung weiter anzupassen oder Schritte zur programmgesteuerten Installation von Anwendungen und zur anschließenden Ausführung benutzerdefinierter Anwendungen zu senden. Bei beiden Optionen zur Clustererstellung können Sie wählen, ob Sie AWS Glue als Ihren SQL Spark-Metastore verwenden möchten. Weitere Informationen finden Sie unter [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).

So starten Sie einen Cluster mit installiertem Spark

1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Cluster erstellen aus, um Schnelle Optionen zu verwenden.
3. Geben Sie einen Clusternamen ein. Ihr Clustername darf die Zeichen <, >, \$, | oder ` (Backtick) nicht enthalten.
4. Wählen Sie unter Softwarekonfiguration die Option Version aus.
5. Wählen Sie für Anwendungen das Spark-Anwendungspaket.
6. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

### Note

Weitere Informationen zum Konfigurieren von Spark beim Erstellen des Clusters finden Sie unter [Konfigurieren von Spark](#).

Um einen Cluster mit installiertem Spark zu starten, verwenden Sie AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl.

```
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

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Um einen Cluster mit installiertem Spark mit dem SDK für Java zu starten

Geben Sie Spark als Anwendung mit `SupportedProductConfig` an, die in `RunJobFlowRequest` verwendet wird.

- Das folgende Beispiel zeigt, wie Sie einen Cluster mit Spark mittels Java erstellen.

```
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 enableddebugging = 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(enableddebugging)
 .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());
}
}
```

# Führen Sie Spark-Anwendungen mit Docker auf Amazon EMR 6.x aus

Mit Amazon EMR 6.0.0 können Spark-Anwendungen Docker-Container verwenden, um ihre Bibliotheksabhängigkeiten zu definieren, anstatt Abhängigkeiten von den einzelnen EC2 Amazon-Instances im Cluster zu installieren. Wenn Sie Spark mit Docker ausführen möchten, müssen Sie zunächst die Docker-Registrierung konfigurieren und beim Senden einer Spark-Anwendung zusätzliche Parameter definieren. Weitere Informationen finden Sie unter [Konfigurieren der Docker-Integration](#).

Wenn die Anwendung eingereicht wird, YARN ruft Docker auf, um das angegebene Docker-Image abzurufen und die Spark-Anwendung in einem Docker-Container auszuführen. Auf diese Weise können Sie Abhängigkeiten einfach definieren und isolieren. Es reduziert die Zeit für das Bootstrapping oder die Vorbereitung von Instances im EMR Amazon-Cluster mit den für die Auftragsausführung benötigten Bibliotheken.

## Überlegungen beim Ausführen von Spark mit Docker

Stellen Sie beim Ausführen von Spark mit Docker sicher, dass die folgenden Voraussetzungen erfüllt sind:

- Das `docker` Paket und die CLI sind nur auf Core- und Task-Nodes installiert.
- Auf Amazon EMR 6.1.0 und höher können Sie Docker alternativ mit den folgenden Befehlen auf einem primären Knoten installieren.
  - ```
sudo yum install -y docker
sudo systemctl start docker
```
- Der `spark-submit` Befehl sollte immer von einer primären Instance auf dem EMR Amazon-Cluster aus ausgeführt werden.
- Die Docker-Registries, die zur Auflösung von Docker-Images verwendet werden, müssen mithilfe der Klassifizierung API mit dem `container-executor` Klassifizierungsschlüssel definiert werden, um beim Starten des Clusters zusätzliche Parameter zu definieren:
 - `docker.trusted.registries`
 - `docker.privileged-containers.registries`
- Wenn Sie eine Spark-Anwendung in einem Docker-Container ausführen möchten, sind die folgenden Konfigurationsoptionen erforderlich:

- `YARN_CONTAINER_RUNTIME_TYPE=docker`
- `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`
- Wenn Sie Amazon ECR zum Abrufen von Docker-Images verwenden, müssen Sie den Cluster so konfigurieren, dass er sich selbst authentifiziert. Dazu müssen Sie die folgende Konfigurationsoption verwenden:
 - `YARN_CONTAINER_RUNTIME_DOCKER_CONFIG = {CLIENT_DOCKER_CLIENT_CONFIG_ON_PATH} HDFS`
- In Amazon EMR 6.1.0 und höher müssen Sie den aufgelisteten Befehl nicht verwenden, `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}` wenn die ECR auto Authentifizierungsfunktion aktiviert ist.
- Bei jedem Docker-Image, das mit Spark verwendet wird, muss Java im Docker-Image installiert sein.

Weitere Informationen zu den Voraussetzungen finden Sie unter [Konfigurieren der Docker-Integration](#).

Ein Docker-Image erstellen

Docker-Images werden mit einer Docker-Datei erstellt, in der die Pakete und die Konfiguration definiert werden, die in das Image aufgenommen werden sollen. Die folgenden beiden Beispiele verwenden Dockerfiles PySpark und SparkR.

PySpark Docker-Datei

Docker-Images, die aus dieser Dockerfile erstellt wurden, enthalten Python 3 und das NumPy Python-Paket. Dieses Dockerfile verwendet Amazon Linux 2 und Amazon JDK Corretto 8.

```
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"
```

SparkR-Docker-Datei

Docker-Images, die aus dieser Dockerfile erstellt wurden, enthalten R und das Paket. randomForest CRAN Dieses Dockerfile beinhaltet Amazon Linux 2 und Amazon JDK Corretto 8.

```
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')"
```

Weitere Informationen zur Docker-Dateisyntax finden Sie in der [Docker-Datei-Referenzdokumentation](#).

Docker-Images von Amazon verwenden ECR

Amazon Elastic Container Registry (Amazon ECR) ist eine vollständig verwaltete Docker-Container-Registry, die das Speichern, Verwalten und Bereitstellen von Docker-Container-Images vereinfacht. Wenn Sie Amazon verwenden ECR, muss der Cluster so konfiguriert sein, dass er Ihrer Instance von vertraut ECR, und Sie müssen die Authentifizierung konfigurieren, damit der Cluster Docker-Images von Amazon ECR verwenden kann. Weitere Informationen finden Sie unter [Konfiguration für YARN den Zugriff auf Amazon ECR](#).

Um sicherzustellen, dass EMR Amazon-Hosts auf die in Amazon gespeicherten Bilder zugreifen können ECR, muss Ihr Cluster über die Berechtigungen aus der

AmazonEC2ContainerRegistryReadOnly Richtlinie verfügen, die dem Instance-Profil zugeordnet ist. Weitere Informationen finden Sie unter [AmazonEC2ContainerRegistryReadOnly-Richtlinie](#).

In diesem Beispiel muss der Cluster mit der folgenden zusätzlichen Konfiguration erstellt werden, um sicherzustellen, dass die ECR Amazon-Registrierung vertrauenswürdig ist. Ersetzen Sie das *123456789123.dkr.ecr.us-east-1.amazonaws.com* Endpunkt mit Ihrem ECR Amazon-Endpunkt.

```
[
  {
    "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": {}
  }
]
```

Verwendung PySpark mit Amazon ECR

Das folgende Beispiel verwendet das PySpark Dockerfile, das markiert und auf Amazon hochgeladen wird. ECR Nachdem Sie das Dockerfile hochgeladen haben, können Sie den PySpark Job ausführen und auf das Docker-Image von Amazon verweisen. ECR

Nachdem Sie den Cluster gestartet haben, stellen Sie eine Verbindung SSH zu einem Core-Node her und führen Sie die folgenden Befehle aus, um das lokale Docker-Image aus dem Dockerfile-Beispiel zu erstellen. PySpark

Erstellen Sie zunächst ein Verzeichnis und eine Docker-Datei.

```
mkdir pyspark
vi pyspark/Dockerfile
```

Fügen Sie den Inhalt der PySpark Dockerfile ein und führen Sie die folgenden Befehle aus, um ein Docker-Image zu erstellen.

```
sudo docker build -t local/pyspark-example pyspark/
```

Erstellen Sie das `emr-docker-examples` ECR Repository für die Beispiele.

```
aws ecr create-repository --repository-name emr-docker-examples
```

Markieren Sie das lokal erstellte Image und laden Sie es hoch. Ersetzen Sie ECR es `123456789123.dkr.ecr.us-east-1.amazonaws.com` mit Ihrem ECR Endpunkt.

```
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
```

Wird verwendet SSH, um eine Verbindung zum primären Knoten herzustellen und ein Python-Skript mit dem Dateinamen `main.py` vorzubereiten. Fügen Sie den folgenden Inhalt in die `main.py`-Datei ein und speichern Sie sie.

```
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)
```

Um den Job auf Amazon EMR 6.0.0 einzureichen, geben Sie den Namen des Docker-Images an. Definieren Sie die zusätzlichen Konfigurationsparameter, um sicherzustellen, dass die Aufgabenausführung Docker als Laufzeit verwendet. Wenn Sie Amazon verwenden ECR, `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` müssen Sie auf die `config.json` Datei verweisen, die die Anmeldeinformationen enthält, die für die Authentifizierung bei Amazon ECR verwendet wurden.

```
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
```

Wenn Sie bei Amazon EMR 6.1.0 und höher den Job einreichen möchten, geben Sie den Namen des Docker-Images an. Wenn die ECR auto Authentifizierung aktiviert ist, führen Sie den folgenden Befehl aus.

```
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
```

Wenn der Job abgeschlossen ist, notieren Sie sich die YARN Anwendungs-ID und verwenden Sie den folgenden Befehl, um die Ausgabe des PySpark Jobs abzurufen.

```
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]]
```

SparkR mit Amazon verwenden ECR

Das folgende Beispiel verwendet das SparkR Dockerfile, das markiert und in das hochgeladen wird. ECR Sobald das Dockerfile hochgeladen ist, können Sie den SparkR ausführen und auf das Docker-Image von Amazon verweisen. ECR

Nachdem Sie den Cluster gestartet haben, stellen Sie eine Verbindung SSH zu einem Core-Knoten her und führen Sie die folgenden Befehle aus, um das lokale Docker-Image aus dem SparkR Dockerfile-Beispiel zu erstellen.

Erstellen Sie zunächst ein Verzeichnis und die Docker-Datei.

```
mkdir sparkr
vi sparkr/Dockerfile
```

Fügen Sie den Inhalt der SparkR-Docker-Datei ein, und führen Sie die folgenden Befehle aus, um ein Docker-Image zu erstellen.

```
sudo docker build -t local/sparkr-example sparkr/
```

Markieren und laden Sie das lokal erstellte Bild auf Amazon hoch und ersetzen Sie ECR es 123456789123.dkr.ecr.us-east-1.amazonaws.com mit Ihrem ECR Amazon-Endpunkt.

```
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
```

Wird verwendet SSH, um eine Verbindung zum Primärknoten herzustellen und ein R-Skript mit dem Namen vorzubereiten sparkR.R. Fügen Sie den folgenden Inhalt in die sparkR.R-Datei ein:

```
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()
```

Verwenden Sie bei Amazon EMR 6.0.0 den Namen des Docker-Images, um den Job einzureichen. Definieren Sie die zusätzlichen Konfigurationsparameter, um sicherzustellen, dass die Aufgabenausführung Docker als Laufzeit verwendet. Wenn Sie Amazon verwenden ECR,

YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG müssen Sie sich auf die `config.json` Datei beziehen, die die für die Authentifizierung verwendeten Anmeldeinformationen enthält. 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
```

Wenn Sie bei Amazon EMR 6.1.0 und höher den Job einreichen möchten, geben Sie den Namen des Docker-Images an. Wenn die ECR auto Authentifizierung aktiviert ist, führen Sie den folgenden Befehl aus.

```
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
```

Wenn der Job abgeschlossen ist, notieren Sie sich die YARN Anwendungs-ID und verwenden Sie den folgenden Befehl, um die Ausgabe des SparkR-Jobs abzurufen. Dieses Beispiel beinhaltet Tests, um sicherzustellen, dass die randomForest Bibliothek, die installierte Version und die Versionshinweise verfügbar sind.

```
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:

Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL

Mit EMR Amazon-Version 5.8.0 oder höher können Sie Spark so konfigurieren, SQL dass der AWS Glue-Datenkatalog als Metastore verwendet wird. Wir empfehlen diese Konfiguration, wenn Sie einen persistenten Metastore oder einen gemeinsamen Metastore für verschiedene Cluster, Services, Anwendungen oder AWS -Konten benötigen.

AWS Glue ist ein vollständig verwalteter Service zum Extrahieren, Transformieren und Laden (ETL), mit dem Sie Ihre Daten einfach und kostengünstig kategorisieren, bereinigen, anreichern und zuverlässig zwischen verschiedenen Datenspeichern verschieben können. Der AWS Glue Data Catalog bietet ein einheitliches Metadaten-Repository für eine Vielzahl von Datenquellen und Datenformaten, das sich sowohl in Amazon EMR als auch in AmazonRDS, Amazon Redshift, Redshift Spectrum, Athena und jede mit dem Apache Hive Metastore kompatible Anwendung integrieren lässt. AWS Glue-Crawler können automatisch Schemas aus Quelldaten in Amazon S3 ableiten und die zugehörigen Metadaten im Datenkatalog speichern. Weitere Informationen zum Datenkatalog finden Sie unter [Füllen des AWS Glue-Datenkatalogs im AWS Glue-Entwicklerhandbuch](#).

Für AWS Glue fallen separate Gebühren an. Es gibt eine monatliche Gebühr für das Speichern und Zugreifen auf die Metadaten im Datenkatalog, einen Stundensatz, der pro Minute für AWS ETL Glue-Jobs und Crawler-Laufzeit abgerechnet wird, und einen Stundensatz, der pro Minute für jeden bereitgestellten Entwicklungsendpunkt abgerechnet wird. Der Datenkatalog ermöglicht Ihnen die gebührenfreie Speicherung von bis zu einer Million Objekten. Wenn Sie mehr als eine Million Objekte speichern, wird Ihnen USD 1 USD für jeweils 100.000 Objekte über eine Million berechnet. Im Data Catalog ist ein Objekt eine Tabelle, eine Partition oder eine Datenbank. Weitere Informationen finden Sie unter [Glue-Preise](#).

⚠ Important

Wenn Sie vor dem 14. August 2017 Tabellen mit Amazon Athena oder Amazon Redshift Spectrum erstellt haben, werden Datenbanken und Tabellen in einem von Athena verwalteten Katalog gespeichert, der vom Glue-Datenkatalog getrennt ist. AWS Um Amazon EMR in diese Tabellen zu integrieren, müssen Sie ein Upgrade auf den AWS Glue Data Catalog durchführen. Weitere Informationen finden Sie unter [Upgrade auf den AWS Glue-Datenkatalog](#) im Amazon Athena Athena-Benutzerhandbuch.

AWS Glue Data Catalog als Metastore angeben

Sie können den AWS Glue-Datenkatalog als Metastore angeben, indem Sie Amazon AWS Management Console EMR API, AWS CLI, oder verwenden. Wenn Sie das CLI oder verwenden API, verwenden Sie die Konfigurationsklassifizierung für Spark, um den Datenkatalog anzugeben. Darüber hinaus können Sie mit Amazon EMR 5.16.0 und höher die Konfigurationsklassifizierung verwenden, um einen Datenkatalog in einem anderen zu spezifizieren. AWS-Konto Wenn Sie die Konsole verwenden, können Sie den Data Catalog mit den erweiterten Optionen oder den Schnelloptionen angeben.

ℹ Note

Die Option zur Verwendung von AWS Glue Data Catalog ist auch bei Zeppelin verfügbar, da Zeppelin mit Spark-Komponenten installiert ist. SQL

Console

Um AWS Glue Data Catalog mit der neuen Konsole als Spark-Metastore anzugeben

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie EC2 im linken Navigationsbereich unter Amazon EMR on die Option Clusters und dann Create cluster aus.
3. Wählen Sie unter Anwendungspaket die Option Spark oder Benutzerdefiniert aus. Wenn Sie Ihren Cluster anpassen, stellen Sie sicher, dass Sie Zeppelin oder Spark als eine Ihrer Anwendungen auswählen.

4. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Spark-Tabellenmetadaten verwenden.
5. Wählen Sie alle anderen Optionen aus, die für Ihren Cluster gelten.
6. Um Ihren Cluster jetzt zu starten, wählen Sie Cluster erstellen aus.

AWS CLI

Um den AWS Glue-Datenkatalog als Spark-Metastore anzugeben mit dem AWS CLI

Weitere Informationen zur Angabe einer Konfigurationsklassifizierung mithilfe von AWS CLI und Amazon EMR API finden Sie unter [Anwendungen konfigurieren](#).

- Geben Sie den Wert für `hive.metastore.client.factory.class` mit der Klassifizierung `spark-hive-site` an, wie in dem folgenden Beispiel gezeigt:

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

Um einen Datenkatalog in einem anderen AWS Konto anzugeben, fügen Sie die `hive.metastore.glue.catalogid` Eigenschaft hinzu, wie im folgenden Beispiel gezeigt. Ersetzen Sie *acct-id* durch das AWS -Konto des Data Catalog.

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

IAMBerechtigungen

Das EC2 Instanzprofil für einen Cluster muss über IAM Berechtigungen für AWS Glue-Aktionen verfügen. Wenn Sie die Verschlüsselung für AWS Glue Data Catalog-Objekte aktivieren, muss die Rolle außerdem berechtigt sein, die für die Verschlüsselung AWS KMS key verwendeten Daten zu verschlüsseln, zu entschlüsseln und zu generieren.

Berechtigungen für AWS Glue-Aktionen

Wenn Sie das EC2 Standard-Instance-Profil für Amazon verwendenEMR, ist keine Aktion erforderlich. Die `AmazonElasticMapReduceforEC2Role` verwaltete Richtlinie, die an die angehängt ist, `EMR_EC2_DefaultRole` ermöglicht alle erforderlichen AWS Glue-Aktionen. Wenn Sie jedoch ein benutzerdefiniertes EC2 Instanzprofil und Berechtigungen angeben, müssen Sie die entsprechenden AWS Glue-Aktionen konfigurieren. Verwenden Sie hierzu die verwaltete Richtlinie `AmazonElasticMapReduceforEC2Role` als Ausgangspunkt. Weitere Informationen finden Sie unter [Servicerolle für EC2 Cluster-Instances \(EC2Instance-Profil\)](#) im Amazon EMR Management Guide.

Berechtigungen zum Verschlüsseln und Entschlüsseln von AWS Glue Data Catalog

Ihr Instance-Profil benötigt die Erlaubnis, Daten mithilfe Ihres Schlüssels zu verschlüsseln und zu entschlüsseln. Sie müssen diese Berechtigungen nicht konfigurieren, wenn die beiden folgenden Aussagen zutreffen:

- Sie aktivieren die Verschlüsselung für AWS Glue Data Catalog-Objekte mithilfe von verwalteten Schlüsseln für AWS Glue.
- Sie verwenden einen Cluster, der sich im selben Verzeichnis befindet AWS-Konto wie der AWS Glue-Datenkatalog.

Andernfalls müssen Sie der Berechtigungsrichtlinie, die Ihrem EC2 Instanzprofil beigefügt ist, die folgende Erklärung hinzufügen.

```
[
  {
    "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"
      }
    ]
  }
]

```

Weitere Informationen zur Verschlüsselung von AWS Glue Data Catalog finden Sie unter [Verschlüsselung Ihres Datenkatalogs](#) im AWS Glue Developer Guide.

Ressourcenbasierte Berechtigungen

Wenn Sie AWS Glue in Verbindung mit Hive, Spark oder Presto in Amazon EMR verwenden, unterstützt AWS Glue ressourcenbasierte Richtlinien zur Steuerung des Zugriffs auf Datenkatalogressourcen. Zu diesen Ressourcen gehören Datenbanken, Tabellen, Verbindungen und benutzerdefinierte Funktionen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für AWS Glue](#) im AWS -Glue-Entwicklerhandbuch.

Wenn Sie ressourcenbasierte Richtlinien verwenden, um den Zugriff auf AWS Glue von Amazon aus zu beschränken, muss der Principal, den Sie in der Berechtigungsrichtlinie angeben, die Rolle sein, die dem EC2 Instance-Profil ARN zugeordnet ist, das bei der Erstellung eines Clusters angegeben wird. Beispielsweise können Sie für eine ressourcenbasierte Richtlinie, die an einen Katalog angehängt ist, die Rolle ARN für die Standard-Servicerolle für Cluster-Instances angeben. EC2 *EMR_EC2_DefaultRole* als der Principal, wobei das im folgenden Beispiel gezeigte Format verwendet wird:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Das Tool *acct-id* kann sich von der AWS Glue-Konto-ID unterscheiden. Dies ermöglicht den Zugriff von EMR Clustern in verschiedenen Konten aus. Sie können mehrere Principals angeben, von denen jeder aus einem anderen Konto stammt.

Überlegungen zur Verwendung von AWS Glue Data Catalog

Beachten Sie die folgenden Punkte, wenn Sie AWS Glue Data Catalog als Metastore mit Spark verwenden:

- Eine Standarddatenbank ohne Speicherort URI führt zu Fehlern beim Erstellen einer Tabelle. Um dieses Problem zu umgehen, verwenden Sie die LOCATION-Klausel, um den Bucket-Speicherort anzugeben, z. B. `s3://EXAMPLE-DOC-BUCKET`, wenn Sie CREATE TABLE verwenden. Alternativ können Sie auch Tabellen in einer anderen Datenbank als der Standarddatenbank erstellen.
- Das Umbenennen von Tabellen in AWS Glue wird nicht unterstützt.
- Wenn Sie eine Hive-Tabelle erstellen, ohne eine LOCATION anzugeben, werden die Tabellendaten an dem in der Eigenschaft `hive.metastore.warehouse.dir` angegebenen Ort gespeichert. Standardmäßig ist dies ein Speicherort in HDFS. Wenn ein anderer Cluster auf die Tabelle zugreifen muss, schlägt er fehl, sofern er nicht über ausreichende Berechtigungen für den Cluster verfügt, der die Tabelle erstellt hat. Da der HDFS Speicher vorübergehend ist, gehen außerdem die Tabellendaten verloren, wenn der Cluster beendet wird, und die Tabelle muss neu erstellt werden. Wir empfehlen, dass Sie LOCATION in Amazon S3 angeben, wenn Sie eine Hive-Tabelle mit AWS Glue erstellen. Alternativ können Sie die `hive-site`-Konfigurationsklassifizierung verwenden, um einen Speicherort in Amazon S3 für `hive.metastore.warehouse.dir` anzugeben, der für alle Hive-Tabellen gilt. Wenn eine Tabelle an einem HDFS Standort erstellt wurde und der Cluster, der sie erstellt hat, noch läuft, können Sie den Tabellenstandort von AWS Glue aus auf Amazon S3 aktualisieren. Weitere Informationen finden Sie unter [Arbeiten mit Tabellen auf der AWS Glue-Konsole](#) im AWS Glue-Entwicklerhandbuch.
- Partitionswerte, die Anführungszeichen und Apostrophe enthalten, werden nicht unterstützt, zum Beispiel `PARTITION (owner="Doe 's")`.
- [Spaltenstatistiken](#) werden für emr-5.31.0 und höher unterstützt.
- Die Verwendung der [Hive-Autorisierung](#) wird nicht unterstützt. Als Alternative sollten Sie die Verwendung [ressourcenbasierter Richtlinien für AWS Glue](#) in Betracht ziehen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für den Amazon EMR Access to AWS Glue-Datenkatalog](#).

Konfigurieren von Spark

Sie können [Spark auf Amazon EMR](#) mit Konfigurationsklassifizierungen konfigurieren. Weitere Informationen über Konfigurationsklassifizierungen finden Sie unter [Anwendungen konfigurieren](#).

Zu den Konfigurationsklassifizierungen für Spark bei Amazon EMR gehören:

- **spark** – Setzt die `maximizeResourceAllocation`-Eigenschaft auf „wahr“ oder „falsch“. Wenn der Wert wahr ist, konfiguriert Amazon EMR automatisch `spark-defaults` Eigenschaften auf der Grundlage der Cluster-Hardwarekonfiguration. Weitere Informationen finden Sie unter [Verwenden von maximizeResourceAllocation](#).
- **spark-defaults** – Setzt Werte in der `spark-defaults.conf`-Datei. Weitere Informationen finden Sie unter [Spark-Konfiguration](#) in der Spark-Dokumentation.
- **spark-env** – Setzt Werte in der `spark-env.sh`-Datei. Weitere Informationen finden Sie unter [Umgebungsvariablen](#) in der Spark-Dokumentation.
- **spark-hive-site** – Legt Werte in der `hive-site.xml` für Spark fest.
- **spark-log4j**— (Amazon EMR veröffentlicht 6.7.x und niedriger) Legt Werte in der `log4j.properties` Datei fest. Weitere Informationen finden Sie in der [log4j.properties.template](#)-Datei auf Github.
- **spark-log4j2**— (Amazon EMR veröffentlicht 6.8.0 und höher) Legt Werte in der `log4j2.properties` Datei fest. Weitere Informationen finden Sie in der [log4j2.properties.template](#)-Datei auf Github.
- **spark-metrics** – Setzt Werte in der `metrics.properties`-Datei. Einstellungen und weitere Informationen finden Sie in der [metrics.properties.template](#)-Datei auf GitHub und unter [Metrics](#) in der Spark-Dokumentation.

Note

Wenn Sie Spark-Workloads EMR von einer anderen Plattform zu Amazon migrieren, empfehlen wir Ihnen, Ihre Workloads mit der zu testen, [Von Amazon festgelegte Spark-StandardEinstellungen EMR](#) bevor Sie benutzerdefinierte Konfigurationen hinzufügen. Die meisten Kunden sehen mit unseren StandardEinstellungen eine verbesserte Leistung.

Themen

- [Von Amazon festgelegte Spark-StandardEinstellungen EMR](#)
- [Konfiguration der Spark-Garbage-Collection auf Amazon EMR 6.1.0](#)
- [Verwenden von maximizeResourceAllocation](#)
- [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#)


- [ThriftServer Spark-Umgebungsvariable](#)
- [Ändern von Spark-StandardEinstellungen](#)
- [Migration von Apache Log4j 1.x zu Log4j 2.x](#)

Von Amazon festgelegte Spark-StandardEinstellungen EMR

Die folgende Tabelle zeigt, wie Amazon Standardwerte EMR festlegt `spark-defaults`, die sich auf Anwendungen auswirken.

Von Amazon festgelegte Spark-StandardEinstellungen EMR

| Einstellung | Beschreibung | Standardwert |
|--|--|---|
| <code>spark.executor.memory</code> | Die pro Executor-Prozess zu verwendende Größe des Speichers. Zum Beispiel: 1g, 2g. | Diese Einstellung wird durch die Core- und Aufgaben-Instance-Typen im Cluster bestimmt. |
| <code>spark.executor.cores</code> | Die für jeden Executor zu verwendende Anzahl der Kerne. | Diese Einstellung wird durch die Core- und Aufgaben-Instance-Typen im Cluster bestimmt. |
| <code>spark.dynamicAllocation.enabled</code> | Wenn „wahr“, verwenden Sie die dynamische Ressourcenzuteilung, um die Anzahl der bei einer Anwendung registrierten Executors basierend auf dem Workload nach oben oder unten zu skalieren. | <code>true</code> (mit Amazon EMR 4.4.0 und höher) |
| <code>spark.sql.hive.advancedPartitionPredicatePushdown.enabled</code> | Wenn der Wert „wahr“ ist, ist der erweiterte Pushdown von Partitionsprädikaten in den Hive-Metastore aktiviert. | <code>true</code> |

 **Note**

Der Spark-Shuffle-Service wird automatisch von Amazon EMR konfiguriert.

| Einstellung | Beschreibung | Standardwert |
|--|---|-------------------|
| <code>spark.sql.hive.stringLikePartitionPredicatePushdown.enabled</code> | Verschiebt und filtert <code>startsWith</code> , <code>contains</code> und <code>endsWith</code> nach unten in den Hive-Metastore. <div data-bbox="591 445 1029 1241" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>Note</p> <p>Glue unterstützt das Prädikat-Pushdown für <code>startsWith</code>, <code>contains</code> oder <code>endsWith</code> nicht. Wenn Sie Glue Metastore verwenden und aufgrund des Prädikat-Pushdowns für diese Funktionen auf Fehler stoßen, setzen Sie diese Konfiguration auf <code>false</code>.</p> </div> | <code>true</code> |

Konfiguration der Spark-Garbage-Collection auf Amazon EMR 6.1.0

Das Einrichten von benutzerdefinierten Garbage-Collection-Konfigurationen mit `spark.driver.extraJavaOptions` und `spark.executor.extraJavaOptions` führt dazu, dass der Treiber- oder Executor-Start mit Amazon EMR 6.1 fehlschlägt, weil eine Konfiguration der Garbage-Collection-Konfiguration mit Amazon 6.1.0 in Konflikt steht. EMR Für Amazon EMR 6.1.0 wird die Standardkonfiguration für die Garbage-Collection über `spark.driver.defaultJavaOptions` und `spark.executor.defaultJavaOptions` festgelegt. Diese Konfiguration gilt nur für Amazon EMR 6.1.0. JVMOptionen, die nichts mit der Garbage-Collection zu tun haben, wie z. B. die für die Konfiguration von logging (`-verbose: class`),

können trotzdem verwendet werden. `extraJavaOptions` Weitere Informationen finden Sie unter [Spark-Anwendungs-Eigenschaften](#).

Verwenden von `maximizeResourceAllocation`

Um Ihre Executors so zu konfigurieren, dass sie auf jedem Knoten in einem Cluster die maximal möglichen Ressourcen nutzen, setzen Sie in Ihrer `spark`-Konfigurationsklassifizierung `maximizeResourceAllocation` auf `true`. Das `maximizeResourceAllocation` ist spezifisch für Amazon EMR. Wenn Sie diese Option aktivieren `maximizeResourceAllocation`, EMR berechnet Amazon die maximalen Rechen- und Speicherressourcen, die für einen Executor auf einer Instance in der Core-Instance-Gruppe verfügbar sind. Anschließend werden die entsprechenden `spark-defaults`-Einstellungen auf der Grundlage der berechneten Höchstwerte festgelegt.

Amazon EMR berechnet die maximalen Rechen- und Speicherressourcen, die für einen Executor verfügbar sind, basierend auf einem Instance-Typ aus der Core-Instance-Flotte. Da jede Instance-Flotte unterschiedliche Instance-Typen und Größen innerhalb einer Flotte haben kann, ist die von Amazon EMR verwendete Executor-Konfiguration möglicherweise nicht die beste für Ihre Cluster. Daher empfehlen wir nicht, die Standardeinstellungen zu verwenden, wenn Sie die maximale Ressourcenzuweisung verwenden. Konfigurieren Sie benutzerdefinierte Einstellungen für Ihre Instance-Flottencluster.

Note

Sie sollten die `maximizeResourceAllocation` Option nicht für Cluster mit anderen verteilten Anwendungen wie verwenden HBase. Amazon EMR verwendet benutzerdefinierte YARN Konfigurationen für verteilte Anwendungen, was zu Konflikten mit Spark-Anwendungen führen `maximizeResourceAllocation` und zum Ausfall führen kann.

Im Folgenden finden Sie ein Beispiel für eine Spark-Konfigurationsklassifizierung mit der `maximizeResourceAllocation`-Einstellung auf `true`.

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```

]

Einstellungen, die in **spark-defaults** konfiguriert sind, wenn **maximizeResourceAllocation** aktiviert ist

| Einstellung | Beschreibung | Wert |
|---------------------------|--|---|
| spark.default.parallelism | Die Standardanzahl von Partitionen wird bei Transformationen wie Join und Parallelize RDDs zurückgegeben reduceByKey, wenn sie nicht vom Benutzer festgelegt wurden. | Zweimal so viele CPU Kerne stehen Containern zur Verfügung. YARN |
| spark.driver.memory | Speichermenge, die für den Treiberprozess verwendet werden soll, d. h. wo er initialisiert SparkContext wird. (zum Beispiel 1 g, 2 g). | Die Einstellung wird auf der Grundlage der Instance-Typen im Cluster konfiguriert. Da die Spark-Treiberanwendung jedoch entweder auf der primären oder auf einer der Kerninstanzen ausgeführt werden kann (z. B. im YARN Client- bzw. Cluster-Modus), wird dies auf der Grundlage des kleineren Instanztyps in diesen beiden Instanzgruppen festgelegt. |
| spark.executor.memory | Die pro Executor-Prozess zu verwendende Größe des Speichers (beispielsweise 1g, 2g). | Die Einstellung wird auf der Grundlage der Core- und der Aufgaben-Instance-Typen im Cluster konfiguriert. |
| spark.executor.cores | Die für jeden Executor zu verwendende Anzahl der Kerne. | Die Einstellung wird auf der Grundlage der Core- und der Aufgaben-Instance-Typen im Cluster konfiguriert. |

| Einstellung | Beschreibung | Wert |
|---------------------------------------|--------------------------|--|
| <code>spark.executor.instances</code> | Die Anzahl der Executor. | Die Einstellung wird auf der Grundlage der Core- und der Aufgaben-Instance-Typen im Cluster konfiguriert. Festgelegt, wenn <code>spark.dynamicAllocation.enabled</code> nicht gleichzeitig explizit auf "true" festgelegt wurde. |

Verhalten der Außerbetriebnahme von Knoten konfigurieren

Mit der EMR Amazon-Version 5.9.0 und höher EMR enthält Spark on Amazon eine Reihe von Funktionen, die sicherstellen, dass Spark die Knotenbeendigung aufgrund einer manuellen Größenänderung oder einer automatischen Skalierungsrichtlinienanforderung ordnungsgemäß verarbeitet. Amazon EMR implementiert in Spark einen Deny-Listing-Mechanismus, der auf dem Mechanismus YARN zur Außerbetriebnahme aufbaut. Dieser Mechanismus hilft sicherzustellen, dass keine neuen Aufgaben in einem Knoten geplant werden, der außer Betrieb genommen wird. Gleichzeitig werden bereits begonnene Aufgaben abgeschlossen. Darüber hinaus gibt es Funktionen, mit deren Hilfe Spark-Aufgaben schneller wiederhergestellt werden, wenn Shuffle-Blöcke bei der Beendigung eines Knotens verloren gehen. Der Neuberechnungsprozess wird früher ausgelöst und ist optimiert, um eine schnellere Neuberechnung mit weniger Stufenwiederholungen zu ermöglichen. Außerdem kann vermieden werden, dass Aufträge infolge von fehlenden Shuffle-Blöcken fehlschlagen.

Important

Die `spark.decommissioning.timeout.threshold` Einstellung wurde in EMR Amazon-Version 5.11.0 hinzugefügt, um die Spark-Resilienz bei der Verwendung von Spot-Instances zu verbessern. Wenn in früheren Versionen ein Knoten eine Spot Instance verwendet und die Instance aufgrund des Bid-Preises beendet wird, kann Spark die Beendigung möglicherweise nicht zuverlässig abschließen. Aufträge können fehlschlagen und Shuffle-Neuberechnungen können eine geraume Zeit in Anspruch nehmen. Aus diesem Grund empfehlen wir die Nutzung von Version 5.11.0 oder höher, wenn Sie Spot Instances verwenden.

Einstellungen für die Außerbetriebnahme von Spark Node

| Einstellung | Beschreibung | Standardwert |
|--|---|-------------------|
| <code>spark.blacklist.decommissioning.enabled</code> | Wenn diese Option auf <code>true</code> gesetzt ist, listet Spark Deny Knoten auf, die sich im <code>decommissioning</code> Status befinden. YARN Spark plant keine neuen Aufgaben auf Executors aus, die auf dem betreffenden Knoten ausgeführt werden. Bereits begonnene Aufgaben werden abgeschlossen. | <code>true</code> |
| <code>spark.blacklist.decommissioning.timeout</code> | Die Zeitspanne, wie lange ein Knoten im Status <code>decommissioning</code> gesperrt wird. Standardmäßig ist dieser Wert auf eine Stunde festgelegt. Dies ist auch die Standardeinstellung für <code>yarn.resourcemanager.decommissioning.timeout</code> . Um sicherzustellen, dass ein Knoten über den gesamten Außerbetriebnahme-Zeitraum gesperrt wird, setzen Sie diesen Wert größer oder gleich <code>yarn.resourcemanager.decommissioning.timeout</code> . Nach Ablauf des Zeitlimits für die Außerbetriebnahme wechselt der Knoten in einen <code>decommiss</code> | 1h |

| Einstellung | Beschreibung | Standardwert |
|-------------|---|--------------|
| | ioned Status, und Amazon EMR kann die Instanz des EC2 Knotens beenden. Wenn noch Aufgaben ausgeführt werden, nachdem die Zeitbeschränkung abgelaufen ist, gehen diese verloren oder werden abgebrochen und auf Executors, die auf anderen Knoten ausgeführt werden, verschoben. | |

| Einstellung | Beschreibung | Standardwert |
|--|--|--------------|
| <code>spark.decommissioning.timeout.threshold</code> | Verfügbar in EMR Amazon-Version 5.11.0 oder höher. In Sekunden angegeben. Wenn ein Knoten in den Status Außerbetriebnahme übergeht und der Host innerhalb eines Zeitraums außer Betrieb genommen wird, der diesem Wert entspricht oder darunter liegt, listet Amazon den Knoten EMR nicht nur auf, sondern bereinigt auch den Host-Status (wie von <code>spark.resourceManager.cleanupExpiredHost</code>), ohne darauf zu warten, dass der Knoten in einen stillgelegten Zustand übergeht. Auf diese Weise kann Spark das Beenden von Spot-Instances besser verarbeiten, weil Spot-Instances innerhalb eines Timeouts von 20 Sekunden außer Betrieb genommen werden, unabhängig von dem Wert von <code>yarn.resourcemanager.decommissioning.timeout</code> , sodass andere Knoten möglicherweise nicht genügend Zeit haben, Shuffle-Dateien zu lesen. | 20s |

| Einstellung | Beschreibung | Standardwert |
|---|---|-------------------|
| <code>spark.resourceManager.cleanupExpiredHost</code> | Wenn die Einstellung auf <code>true</code> gesetzt ist, löscht Spark alle zwischengespeicherten Daten und Shuffle-Blöcke, die in Executors in Knoten gespeichert sind, die sich im Zustand <code>decommissioned</code> befinden. Dadurch wird der Wiederherstellungsprozess beschleunigt. | <code>true</code> |
| <code>spark.stage.attempt.ignoreOnDecommissionFetchFailure</code> | Wenn die Einstellung auf <code>true</code> gesetzt ist, wird verhindert, dass in Spark Stufen und letztendlich Aufträge wegen zu vieler Abrufe von außer Betrieb genommenen Knoten fehlschlagen. Fehler beim Abrufen von Shuffle-Blöcken von einem Knoten im Zustand <code>decommissioned</code> werden nicht zu der maximal zulässigen Anzahl von aufeinander folgenden Abruffehlern hinzugezählt. | <code>true</code> |

ThriftServer Spark-Umgebungsvariable

Spark legt die Hive-Thrift-Server-Port-Umgebungsvariable `HIVE_SERVER2_THRIFT_PORT` auf "10001" fest.

Ändern von Spark-StandardEinstellungen

Sie ändern die Standardwerte in `spark-defaults.conf` mit der Konfigurationsklassifizierung `spark-defaults` oder der Einstellung `maximizeResourceAllocation` in der Konfigurationsklassifizierung `spark`.

Die folgenden Verfahren zeigen, wie Sie Einstellungen mit der CLI OR-Konsole ändern.

Um einen Cluster zu erstellen, bei dem `spark.executor.memory` auf 2g gesetzt ist, verwenden Sie CLI

- Erstellen Sie einen Cluster mit installiertem Spark und `spark.executor.memory` auf 2G gesetzt, indem Sie den folgenden Befehl eingeben, der auf die Datei `myConfig.json` verweist, die in Amazon S3 gespeichert ist.

```
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

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

`myConfig.json`:

```
[  
  {  
    "Classification": "spark-defaults",  
    "Properties": {  
      "spark.executor.memory": "2G"  
    }  
  }  
]
```

Erstellen Sie einen Cluster mithilfe der Konsole mit der Einstellung „spark.executor.memory“ auf „2G“ wie folgt

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie Spark aus.
4. Lassen Sie unter Edit software settings (Softwareeinstellungen bearbeiten) die Option Enter configuration (Konfiguration eingeben) ausgewählt und nehmen Sie folgende Konfiguration vor:

```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. Wählen Sie andere Optionen aus, dann und anschließend Create cluster (Cluster erstellen).

Zum Einstellen maximizeResourceAllocation

- Erstellen Sie einen Cluster, auf dem Spark installiert und auf true maximizeResourceAllocation gesetzt ist AWS CLI, indem Sie die in Amazon S3 gespeicherte Datei referenzierend verwenden. myConfig.json

```
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

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[  
{
```

```
"Classification": "spark",
"Properties": {
  "maximizeResourceAllocation": "true"
}
]
```

Note

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Sie tun dies, indem Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Migration von Apache Log4j 1.x zu Log4j 2.x

[Apache-Spark-Versionen](#) 3.2.x und frühere Versionen verwenden den älteren Apache Log4j 1.x und die `log4j.properties`-Datei, um Log4j in Spark-Prozessen zu konfigurieren. Apache-Spark-Versionen 3.3.0 und höher verwenden Apache Log4j 2.x und die Datei `log4j2.properties`, um Log4j in Spark-Prozessen zu konfigurieren.

Wenn Sie Apache Spark Log4j mit einer EMR Amazon-Version vor 6.8.0 konfiguriert haben, müssen Sie die alte `spark-log4j` Konfigurationsklassifizierung entfernen und zur Konfigurationsklassifizierung und zum Schlüsselformat migrieren, bevor Sie ein Upgrade auf Amazon EMR 6.8.0 oder höher durchführen können. `spark-log4j2` Die `spark-log4j` Legacy-Klassifizierung führt dazu, dass die Clustererstellung in EMR Amazon-Versionen 6.8.0 und höher mit einem `ValidationException` Fehler fehlschlägt. Ein Fehler im Zusammenhang mit der Log4j-Inkompatibilität wird Ihnen nicht in Rechnung gestellt. Sie müssen jedoch die Klassifizierung für die veraltete `spark-log4j`-Konfiguration entfernen, um fortzufahren.

Weitere Informationen zur Migration von Apache Log4j 1.x zu Log4j 2.x finden Sie im [Apache Log4j Migration Benutzerhandbuch](#) und im [Spark Log4j 2 Template](#) auf Github.

Note

Bei Amazon EMR verwendet Apache Spark eine `log4j2.properties` Datei anstelle der im [Apache Log4j Migration Guide](#) beschriebenen `.xml`-Datei. Außerdem empfehlen wir nicht, die `Log4j-1.x-Bridge`-Methode für die Konvertierung nach `Log4j 2.x` zu verwenden.

Die Leistung von Spark optimieren

Amazon EMR bietet mehrere Funktionen zur Leistungsoptimierung für Spark. In diesem Thema werden die einzelnen Optimierungsfunktionen im Detail erläutert.

Weitere Informationen zum Festlegen der Spark-Konfiguration finden Sie unter [Konfigurieren von Spark](#).

Adaptive Abfrageausführung

Die adaptive Abfrageausführung ist ein Framework zur Neuoptimierung von Abfrageplänen auf der Grundlage von Laufzeitstatistiken. Ab Amazon EMR 5.30.0 sind die folgenden Optimierungen der adaptiven Abfrageausführung von Apache Spark 3 auf Apache Amazon EMR Runtime for Spark 2 verfügbar.

- Adaptive Join-Konvertierung
- Adaptives Zusammenführen von Shuffle-Partitionen

Adaptive Join-Konvertierung

Die adaptive Join-Konvertierung verbessert die Abfrageleistung, indem `sort-merge-join` Operationen in Operationen umgewandelt werden, `broadcast-hash-joins` die auf der Laufzeitgröße der Abfragephasen basieren. `Broadcast-hash-joins` schneidet tendenziell besser ab, wenn eine Seite des Joins klein genug ist, um seine Ausgabe effizient an alle Executoren zu verteilen. Dadurch entfällt die Notwendigkeit, den Austausch zu mischen und beide Seiten des Joins zu sortieren. Die adaptive Join-Konvertierung erweitert das Spektrum der Fälle, in denen Spark automatisch ausgeführt wird.

`broadcast-hash-joins`

Dieses Feature ist standardmäßig aktiviert. Sie kann deaktiviert werden, indem `spark.sql.adaptive.enabled` auf `false` gesetzt wird, wodurch auch der

adaptive Abfrageausführungsrahmen deaktiviert wird. Spark entscheidet, a in a sort-merge-join zu konvertieren, broadcast-hash-join wenn die Laufzeitgrößenstatistik einer der Join-Seiten den Standardwert von 10.485.760 Byte (10 MiB) nicht überschreitet `spark.sql.autoBroadcastJoinThreshold`.

Adaptives Zusammenführen von Shuffle-Partitionen

Das adaptive Zusammenführen von Shuffle-Partitionen verbessert die Abfrageleistung, indem kleine zusammenhängende Shuffle-Partitionen zusammengeführt werden, um den Mehraufwand zu vermeiden, der durch zu viele kleine Aufgaben entsteht. Auf diese Weise können Sie eine höhere Anzahl von anfänglichen Shuffle-Partitionen konfigurieren, die dann zur Laufzeit auf eine bestimmte Größe reduziert werden, wodurch sich die Wahrscheinlichkeit erhöht, dass Shuffle-Partitionen gleichmäßiger verteilt werden.

Dieses Feature ist standardmäßig aktiviert, sofern nicht ausdrücklich `spark.sql.shuffle.partitions` festgelegt ist. Sie kann aktiviert werden, indem Sie `spark.sql.adaptive.coalescePartitions.enabled` auf `true` einstellen. Sowohl die anfängliche Anzahl der Shuffle-Partitionen als auch die Größe der Zielpartition können mit den Eigenschaften `spark.sql.adaptive.coalescePartitions.minPartitionNum` und `spark.sql.adaptive.advisoryPartitionSizeInBytes` eingestellt werden. In der folgenden Tabelle finden Sie weitere Informationen zu den zugehörigen Spark-Eigenschaften für dieses Feature.

Adaptive Koaleszenzpartitioneigenschaften von Spark

| Eigenschaft | Standardwert | Beschreibung |
|--|--|---|
| <code>spark.sql.adaptive.coalescePartitions.enabled</code> | „Wahr“, sofern <code>spark.sql.shuffle.partitions</code> nicht explizit festgelegt ist | Wenn „Wahr“ und <code>spark.sql.adaptive.enabled</code> den Wert „Wahr“ hat, fügt Spark zusammenhängende Shuffle-Partitionen entsprechend der Zielgröße (angegeben durch <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>) zusammen, um zu viele kleine Aufgaben zu vermeiden. |

| Eigenschaft | Standardwert | Beschreibung |
|--|--------------|---|
| <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> | 64 MB | Die empfohlene Größe der Shuffle-Partition beim Zusammenführen. Diese Konfiguration wirkt sich nur aus, wenn <code>spark.sql.adaptive.enabled</code> und <code>spark.sql.adaptive.coalescePartitions.enabled</code> beide gleichzeitig <code>true</code> sind. |
| <code>spark.sql.adaptive.coalescePartitions.minPartitionNum</code> | 25 | Die Mindestanzahl der Shuffle-Partitionen nach dem Zusammenführen. Diese Konfiguration wirkt sich nur aus, wenn <code>spark.sql.adaptive.enabled</code> und <code>spark.sql.adaptive.coalescePartitions.enabled</code> beide gleichzeitig <code>true</code> sind. |
| <code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code> | 1000 | Die anfängliche Anzahl von Shuffle-Partitionen vor dem Zusammenführen. Diese Konfiguration wirkt sich nur aus, wenn <code>spark.sql.adaptive.enabled</code> und <code>spark.sql.adaptive.coalescePartitions.enabled</code> beide gleichzeitig <code>true</code> sind. |

Dynamische Partitionsbereinigung

Die dynamische Partitionsbereinigung verbessert die Auftragsleistung durch eine sorgfältigere Auswahl der spezifischen Partitionen in einer Tabelle, die für eine bestimmte Abfrage gelesen und verarbeitet werden müssen. Indem Sie die Datenmenge reduzieren, die für die Ausführung eines Auftrags gelesen und verarbeitet werden muss, können Sie viel Zeit sparen. Mit Amazon EMR 5.26.0 ist diese Funktion standardmäßig aktiviert. Mit Amazon EMR 5.24.0 und 5.25.0 können Sie diese Funktion aktivieren, indem Sie die Spark-Eigenschaft in Spark oder beim Erstellen `spark.sql.dynamicPartitionPruning.enabled` von Clustern festlegen.

Dynamisches Löschen von Partitionen in Spark, Partitioneigenschaften

| Eigenschaft | Standardwert | Beschreibung |
|--|-------------------|--|
| <code>spark.sql.dynamicPartitionPruning.enabled</code> | <code>true</code> | Wenn der Wert wahr ist, aktivieren Sie das dynamische Bereinigen von Partitionen. |
| <code>spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse</code> | <code>true</code> | Wenn <code>true</code> , führt Spark vor der Ausführung der Abfrage eine Defensivprüfung durch, um sicherzustellen, dass die Wiederverwendung von Broadcast-Austauschen in dynamischen Bereinigungsfiltern nicht durch spätere Vorbereitungsregeln, wie z. B. benutzerdefinierte Spaltenregeln, beeinträchtigt wird. Wenn die Wiederverwendung nicht funktioniert und diese Konfiguration <code>true</code> ist, entfernt Spark die betroffenen dynamischen Bereinigungsfilter, um Leistungs- und Korrektheitsprobleme zu vermeiden. Korrektheitsprobleme können |

| Eigenschaft | Standardwert | Beschreibung |
|-------------|--------------|--|
| | | auftreten, wenn der Broadcast-Austausch des dynamischen Bereinigungsfilters zu unterschiedlichen, inkonsistenten Ergebnissen führt, als der Broadcast-Austausch des entsprechenden Join-Vorgangs. Das Einstellen dieser Konfiguration auf <code>false</code> sollte mit Vorsicht erfolgen. Dadurch können Szenarien umgangen werden, z. B. wenn die Wiederverwendung durch benutzerdefinierte Spaltenregeln gestört wird. Wenn Adaptive Abfrageausführung aktiviert ist, wird die Wiederverwendung von Broadcasts immer erzwungen. |

Diese Optimierung verbessert die vorhandenen Funktionen von Spark 2.4.2, das nur die Weitergabe statischer Prädikate unterstützt, um diese zu geplanten Zeiten aufzulösen.

Im Folgenden finden Sie Beispiele für die Weitergabe statischer Prädikate in Spark 2.4.2.

```
partition_col = 5

partition_col IN (1,3,5)

partition_col between 1 and 3

partition_col = 1 + 3
```

Die dynamische Partitionsbereinigung ermöglicht es der Spark-Engine, während der Laufzeit dynamisch abzuleiten, welche Partitionen gelesen werden müssen und welche problemlos eliminiert werden können. Die folgende Abfrage beinhaltet beispielsweise zwei Tabellen: die Tabelle

`store_sales`, die den Gesamtumsatz aller Geschäfte enthält und nach Regionen partitioniert ist, sowie die Tabelle `store_regions`, die für die einzelnen Länder eine Zuweisung nach Regionen enthält. Die Tabellen enthalten Daten für Geschäfte, die auf der ganzen Welt verteilt sind, allerdings benötigen wir nur die Daten für Nordamerika.

```
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'
```

Ohne die dynamische Partitionsbereinigung, liest diese Anfrage alle Regionen, bevor Sie die Untergruppe an Regionen herausfiltert, die mit den Ergebnissen der Unterabfrage übereinstimmen. Mit der dynamischen Partitionsbereinigung liest und verarbeitet diese Abfrage nur die Partitionen für die Regionen, die in der Unterabfrage zurückgegeben wurden. Da weniger Daten im Speicher gelesen und weniger Datensätze verarbeitet werden müssen, spart dies Zeit und Ressourcen.

Abflachen skalarer Unterabfragen

Diese Optimierung verbessert die Leistung von Abfragen, die in der gleichen Tabelle skalare Unterabfragen ausführen. Mit Amazon EMR 5.26.0 ist diese Funktion standardmäßig aktiviert. Mit Amazon EMR 5.24.0 und 5.25.0 können Sie es aktivieren, indem Sie die Spark-Eigenschaft `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` von Clustern festlegen. Wenn diese Eigenschaft auf „true“ festgelegt ist, flacht der Abfrageoptimierer aggregierte skalare Unterabfragen, die, wenn möglich, dieselbe Relation verwenden, ab. Die skalaren Unterabfragen werden abgeflacht, indem alle in der Unterabfrage vorhandenen Prädikate an die Aggregationsfunktionen weitergegeben werden. Im Anschluss daran wird pro Relation eine Aggregation mit allen Aggregatfunktionen ausgeführt.

Im Folgenden finden Sie ein Beispiel für eine Abfrage, die von dieser Optimierung profitiert.

```
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
```

Die Optimierung schreibt die vorherige Abfrage folgendermaßen um:

```
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);
```

Beachten Sie, dass die umgeschriebene Abfrage die Studententabelle nur einmal liest und die Prädikate der drei Unterabfragen per Push-Verfahren in die avg-Funktion weitergegeben werden.

DISTINCTvorher INTERSECT

Diese Optimierung optimiert Verknüpfungen bei der Verwendung von INTERSECT. Mit Amazon EMR 5.26.0 ist diese Funktion standardmäßig aktiviert. Mit Amazon EMR 5.24.0 und 5.25.0 können Sie es aktivieren, indem Sie die Spark-Eigenschaft in Spark oder beim Erstellen `spark.sql.optimizer.distinctBeforeIntersect.enabled` von Clustern festlegen. Abfragen, die verwenden, INTERSECT werden automatisch so konvertiert, dass sie einen Left-Semi-Join verwenden. Wenn diese Eigenschaft auf „true“ gesetzt ist, leitet der Abfrageoptimierer den DISTINCT Operator an die untergeordneten Elemente von weiter, INTERSECT wenn er feststellt, dass der DISTINCT Operator die Links-Halbverknüpfung zu einem anstelle von a BroadcastHashJoin machen kann. SortMergeJoin

Im Folgenden finden Sie ein Beispiel für eine Abfrage, die von dieser Optimierung profitiert.

```
(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)
```

Wenn die Eigenschaft `spark.sql.optimizer.distinctBeforeIntersect.enabled`, nicht aktiviert ist, wird die Abfrage folgendermaßen neu geschrieben.

```
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
```

Wenn die Eigenschaft `spark.sql.optimizer.distinctBeforeIntersect.enabled`, nicht aktiviert ist, wird die Abfrage folgendermaßen neu geschrieben.

```
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
```

Bloomfilter für Joins

Durch diese Optimierung kann die Leistung einiger Joins verbessert werden, da eine Seite eines Joins mit einem [Bloomfilter](#) vorgefiltert wird, der aus den Werten von der anderen Seite des Joins generiert wird. Mit Amazon EMR 5.26.0 ist diese Funktion standardmäßig aktiviert. Mit Amazon EMR 5.25.0 können Sie diese Funktion aktivieren, indem Sie die Spark-Eigenschaft in Spark oder beim Erstellen `true` von Clustern `spark.sql.bloomFilterJoin.enabled` auf einstellen.

Im Folgenden finden Sie eine Beispielabfrage, für die die Anwendung eines Bloomfilters geeignet wäre.

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```

Wenn diese Funktion aktiviert ist, wird der Bloomfilter aus allen Artikel-IDs erstellt, deren Kategorie sich innerhalb des Kategoriensatzes befindet, der abgefragt wird. Beim Scannen der Tabelle SALES wird der Bloomfilter verwendet, um zu ermitteln, welche Verkäufe für Artikel gelten, die sich definitiv nicht im vom Bloomfilter definierten Satz befinden. So können diese ermittelten Verkäufe so früh wie möglich herausgefiltert werden.

Optimierte Join-Neuanordnung

Durch diese Optimierung kann die Abfrageleistung verbessert werden, indem bei Tabellen mit Filtern die Joins neu angeordnet werden. Mit Amazon EMR 5.26.0 ist diese Funktion standardmäßig aktiviert. Mit Amazon EMR 5.25.0 können Sie diese Funktion aktivieren, indem Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` auf `true` setzen. Das Standardverhalten von Spark ist, dass Tabellen von links nach rechts verknüpft werden,

wie in der Abfrage aufgeführt. Dabei kann die Chance verpasst werden, zuerst kleinere Joins mit Filtern auszuführen, was später teureren Joins zugute kommen würde.

Bei der Beispielabfrage unten werden alle zurückgegebenen Artikel aus allen Läden in einem Land ausgewertet. Ohne optimierte Join-Neuanordnung verknüpft Spark zuerst die zwei großen Tabellen `store_sales` und `store_returns` und verknüpft sie dann mit `store` und schließlich mit `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'
```

Bei der optimierten Join-Neuanordnung verknüpft Spark zuerst `store_sales` mit `store`, da `store` über einen Filter verfügt und kleiner ist als `store_returns` und `broadcastable`. Danach führt Spark einen Join mit `store_returns` und schließlich mit `item` durch. Wenn `item` über einen Filter verfügen würde und sendungsfähig wäre, würde es sich ebenfalls für die Neuordnung eignen, was dazu führen würde, dass `store_sales` mit `store`, dann mit `item` und schließlich mit `store_returns` verknüpft würde.

Caching von Ergebnisfragmenten in Spark

Amazon EMR 6.6.0 und höher enthalten die optionale Spark-Funktion zum Zwischenspeichern von Ergebnisfragmenten, die automatisch Ergebnisfragmente zwischenspeichert. Diese Ergebnisfragmente sind Teile von Ergebnissen aus Teilbäumen von Abfragen, die in einem Amazon-S3-Bucket Ihrer Wahl gespeichert sind. Die gespeicherten Fragmente der Abfrageergebnisse werden bei nachfolgenden Abfrageausführungen wiederverwendet, was zu schnelleren Abfragen führt.

Result Fragment Caching analysiert Ihre SQL Spark-Abfragen und speichert geeignete Ergebnisfragmente an Ihrem angegebenen S3-Speicherort im Cache. Bei nachfolgenden Abfrägeläufen werden die verwendbaren Fragmente der Abfrageergebnisse automatisch erkannt und von S3 abgerufen. Das Caching von Ergebnisfragmenten unterscheidet sich vom Zwischenspeichern von Ergebnismengen, bei dem nachfolgende Abfragen exakt mit der ursprünglichen Abfrage übereinstimmen müssen, um Ergebnisse aus dem Cache zurückzugeben. Wenn es für Abfragen verwendet wird, die wiederholt auf eine statische Teilmenge Ihrer Daten abzielen, beschleunigt das Caching von Ergebnisfragmenten die Leistung erheblich.

Stellen Sie sich die folgende Abfrage vor, die Bestellungen bis zum Jahr 2022 zählt:

```
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
```

Im Laufe der Zeit muss diese Abfrage täglich ausgeführt werden, um den Gesamtumsatz des Jahres zu melden. Ohne das Caching von Ergebnisfragmenten müssen die Ergebnisse für alle Tage des Jahres täglich neu berechnet werden. Die Abfrage wird mit der Zeit langsamer und am Jahresende, wenn die Ergebnisse aller 365 Tage neu berechnet werden müssen, am langsamsten.

Wenn Sie das Caching von Ergebnisfragmenten aktivieren, verwenden Sie Ergebnisse für alle vorherigen Tage des Jahres aus dem Cache. An jedem Tag muss das Feature nur die Ergebnisse eines Tages neu berechnen. Nachdem das Feature das Ergebnisfragment berechnet hat, speichert das Feature das Fragment im Cache. Dadurch sind die Cache-aktivierten Abfragezeiten schnell und sie bleiben bei jeder nachfolgenden Abfrage konstant.

Caching von Ergebnisfragmenten in Spark aktivieren

Zum Aktivieren von Caching von Ergebnisfragmenten in Spark führen Sie die folgenden Schritte durch:

1. Erstellen Sie einen Cache-Bucket in Amazon S3 und autorisieren Sie den Lese-/Schreibzugriff für. EMRFS Weitere Informationen finden Sie unter [Autorisieren des Zugriffs auf EMRFS Daten in Amazon S3](#).
2. Stellen Sie die Amazon EMR Spark-Konfiguration ein, um die Funktion zu aktivieren.

```
spark.subResultCache.enabled = true  
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. Aktivieren Sie das S3-Lebenszyklusmanagement für den Bucket, um Cache-Dateien automatisch zu bereinigen.
4. Optional können Sie die `maxBufferSize` Eigenschaften `reductionRationThreshold` und konfigurieren, um die Funktion weiter zu optimieren.


```
spark.sql.subResultCache.reductionRatioThreshold  
spark.sql.subResultCache.maxBufferSize
```

Einige Hinweise zur Verwendung von Caching von Ergebnisfragmenten

Die Kosteneinsparungen, wenn Sie Ergebnisse verwenden, die bereits in Amazon S3 zwischengespeichert wurden, anstatt sie neu zu berechnen, steigen mit der Häufigkeit, mit der dieselben zwischengespeicherten Ergebnisse verwendet werden können. Abfragen mit umfangreichen Tabellenscans, gefolgt von Filtern oder Hash-Aggregationen, die die Ergebnisgröße um den Faktor 8 reduzieren (d. h. ein Verhältnis von mindestens 8:1 bei Eingabegröße zu Ergebnissen), profitieren am meisten von diesem Feature. Je größer das Reduktionsverhältnis zwischen Eingabe und Ergebnissen ist, desto größer ist der Kostenvorteil. Abfragen mit kleineren Reduktionsraten, die jedoch teure Rechenschritte zwischen dem Tabellenscan und Filtern oder Aggregationen enthalten, werden ebenfalls davon profitieren, sofern die Kosten für die Erstellung der Ergebnisse höher sind als die Kosten für das Abrufen der Ergebnisse aus Amazon S3. Standardmäßig wird das Caching von Ergebnisfragmenten nur wirksam, wenn erkannt wird, dass das Reduktionsverhältnis mindestens 8:1 beträgt.

Wenn Ihre Abfragen zwischengespeicherte Ergebnisse wiederholt wiederverwenden, sind die Vorteile dieses Features am größten. Rollende und inkrementelle Fensterabfragen sind gute Beispiele. Beispielsweise müsste bei einer 30-tägigen rollenden Fenster-Abfrage, die bereits 29 Tage lang ausgeführt wurde, nur 1/30 der Zieldaten aus der ursprünglichen Eingabequelle abgerufen werden, und es würden zwischengespeicherte Ergebnisfragmente für die 29 vorherigen Tage verwendet. Eine inkrementelle Fensterabfrage wäre sogar noch besser, da der Beginn des Fensters unverändert bleibt: Bei jedem Aufruf der Abfrage muss ein kleinerer Prozentsatz der Verarbeitung aus der Eingabequelle gelesen werden.

Im Folgenden finden Sie weitere Überlegungen zur Verwendung von Caching von Ergebnisfragmenten:

- Abfragen, die nicht auf dieselben Daten mit denselben Abfragefragmenten abzielen, haben eine niedrige Cache-Trefferrate und profitieren daher nicht von diesem Feature.
- Abfragen mit niedrigen Reduktionsraten, die keine teuren Rechenschritte enthalten, führen zu zwischengespeicherten Ergebnissen, deren Lesen ungefähr so teuer ist wie die ursprüngliche Verarbeitung.

- Die erste Abfrage weist aufgrund der Kosten für das Schreiben in den Cache immer eine geringfügige Regression auf.
- Das Feature zum Caching von Ergebnisfragmenten funktioniert ausschließlich mit Parquet-Dateien. Andere Dateiformate werden nicht unterstützt.
- Das Feature Caching von Ergebnisfragmenten für das Zwischenspeichern von Ergebnisfragmenten versucht nur, Scans mit Datei-Split-Größen von 128 MB oder mehr zwischenzuspeichern. Bei der Standardkonfiguration von Spark ist das Caching von Ergebnisfragmenten deaktiviert, wenn die Scangröße (Gesamtgröße aller gescannten Dateien) geteilt durch die Anzahl der Executor-Cores weniger als 128 MB beträgt. Wenn eine der unten aufgeführten Spark-Konfigurationen festgelegt ist, beträgt die Größe der aufgeteilten Datei:

```
min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
```

- `spark.sql.leafNodeDefaultParallelismus` (Standardwert ist `spark.default.parallelism`)
- `spark.sql.files.minPartitionNum` (Der Standardwert ist `spark.sql.leafNodeDefaultParallelität`)
- `spark.sql.files.openCostInByte`
- `spark.sql.dateien.maxPartitionBytes`
- Die Funktion zum Zwischenspeichern von Ergebnisfragmenten speichert die RDD Partitionsgranularität im Cache. Das zuvor beschriebene Reduktionsverhältnis, das standardmäßig 8:1 beträgt, wird pro Partition bewertet. RDD Bei Workloads, bei denen das Verhältnis pro RDD Reduzierung sowohl größer als auch kleiner als 8:1 ist, kann es zu geringeren Leistungsvorteilen kommen als bei Workloads mit einem Verhältnis pro RDD Reduzierung, das durchweg unter 8:1 liegt.
- Die Funktion zum Zwischenspeichern von Ergebnisfragmenten verwendet standardmäßig einen Schreibpuffer von 16 MB für jede RDD Partition, die zwischengespeichert wird. Wenn mehr als 16 MB pro RDD Partition zwischengespeichert werden, können die Kosten für die Feststellung, dass kein Schreibvorgang möglich ist, zu Leistungseinbußen führen.
- Standardmäßig versucht Result Fragment Caching zwar nicht, RDD Partitionsergebnisse mit einem Reduktionsverhältnis von weniger als 8:1 zwischenzuspeichern und seinen Schreibpuffer auf 16 MB begrenzt, aber beide Werte können über die folgenden Konfigurationen eingestellt werden:

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)  
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- Mehrere Cluster, die dieselbe EMR Amazon-Version verwenden, können sich denselben Cache-Speicherort teilen. Um die Richtigkeit der Ergebnisse sicherzustellen, verwendet Result

Fragment Caching keine Cache-Ergebnisse, die von verschiedenen Versionen von Amazon EMR geschrieben wurden.

- Das Zwischenspeichern von Ergebnisfragmenten wird automatisch für Spark Streaming-Anwendungsfälle deaktiviert oder wenn RecordServer Apache Ranger oder AWS Lake Formation verwendet wird.
- Das Ergebnis verwendet im Fragment-Cache Lese/Schreibvorgänge EMRFS und Amazon S3 S3-Buckets. CSE/SSES3/Verschlüsselung SSE KMS wird unterstützt.

Verwenden Sie den Nvidia RAPIDS Accelerator für Apache Spark

Mit EMR Amazon-Version 6.2.0 und höher können Sie das [RAPIDS Accelerator for Apache Spark-Plug-In](#) von Nvidia verwenden, um Spark mithilfe von EC2 Graphics Processing Unit (GPU) - Instance-Typen zu beschleunigen. RAPIDS Accelerator GPU beschleunigt Ihre Apache Spark 3.0-Pipelines für die Datenwissenschaft ohne Codeänderungen, beschleunigt die Datenverarbeitung und das Modelltraining und senkt gleichzeitig die Infrastrukturkosten erheblich.

Die folgenden Abschnitte führen Sie durch die Konfiguration Ihres EMR Clusters für die Verwendung des RAPIDS Spark-Plug-ins für Spark.

Instance-Typen auswählen

Um das Nvidia RAPIDS Spark-Plugin für Spark verwenden zu können, müssen die Core- und Task-Instanzgruppen EC2 GPU Instanztypen verwenden, die den [Hardwareanforderungen](#) von Spark entsprechen RAPIDS. Eine vollständige Liste der von Amazon EMR unterstützten GPU Instance-Typen finden Sie unter [Unterstützte Instance-Typen](#) im Amazon EMR Management Guide. Der Instance-Typ für die primäre Instance-Gruppe kann entweder ein Typ GPU oder ein anderer GPU Typ sein, ARM Instance-Typen werden jedoch nicht unterstützt.

Anwendungskonfigurationen für Ihren Cluster einrichten

1. Ermöglichen EMR Sie Amazon, die Plugins auf Ihrem neuen Cluster zu installieren

Zum Installieren von Plugins geben Sie beim Erstellen Ihres Clusters die folgende Konfiguration an:

```
{
  "Classification": "spark",
  "Properties": {
    "enableSparkRapids": "true"
  }
}
```

```
}
}
```

2. YARN zur Verwendung konfigurieren GPU

Einzelheiten zur Verwendung von GPU finden Sie unter [GPU YARN Using GPU on YARN in der Apache Hadoop-Dokumentation](#). Die folgenden Beispiele zeigen YARN Beispielfiguren für Amazon EMR 6.x- und 7.x-Versionen:

Amazon EMR 7.x

YARN Beispielfiguren für 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

YARN-Beispielkonfiguration für 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. Konfigurieren Sie Spark für die Verwendung RAPIDS

Hier sind die erforderlichen Konfigurationen, damit Spark das RAPIDS Plugin verwenden kann:

```

{
  "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-Spark-Bibliothek](#) in der XGBoost Dokumentation ist auch verfügbar, wenn das RAPIDS Spark-Plugin auf Ihrem Cluster aktiviert ist. Sie können die folgende Konfiguration für die Integration in Ihren XGBoost Spark-Job verwenden:

```

{
  "Classification": "spark-defaults",
  "Properties": {
    "spark.submit.pyFiles": "/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
  }
}

```

Weitere Spark-Konfigurationen, mit denen Sie einen GPU -beschleunigten EMR Cluster optimieren können, finden Sie im [Rapids Accelerator for Apache Spark-Tuning-Leitfaden](#) in der NVIDIA.GitHub.io-Dokumentation.

4. YARNCapacity Scheduler konfigurieren

DominantResourceCalculator muss so konfiguriert sein, dass GPU Planung und Isolierung aktiviert sind. Weitere Informationen finden Sie unter [Using GPU on YARN in der](#) Apache Hadoop-Dokumentation.

```
{
  "Classification": "capacity-scheduler",
  "Properties": {
    "yarn.scheduler.capacity.resource-
calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
  }
}
```

5. Erstellen Sie eine JSON Datei, die Ihre Konfigurationen enthält

Sie können eine JSON Datei erstellen, die Ihre Konfiguration enthält, um das RAPIDS Plugin für Ihren Spark-Cluster zu verwenden. Sie stellen die Datei später bereit, wenn Sie Ihren Cluster starten.

Sie können die Datei lokal oder in S3 speichern. Weitere Informationen zur Bereitstellung von Anwendungskonfigurationen für Ihre Cluster finden Sie unter [Anwendungen konfigurieren](#).

Verwenden Sie die folgenden Beispieldateien als Vorlagen, um Ihre eigenen Konfigurationen zu erstellen.

Amazon EMR 7.x

my-configurations.json Beispieldatei für 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 Beispieldatei für 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"
    }
  }
]
```

Fügen Sie eine Bootstrap-Aktion für Ihren Cluster hinzu

Weitere Informationen zur Bereitstellung von Bootstrap-Aktionsskripten bei der Erstellung Ihres Clusters finden Sie unter [Grundlagen der Bootstrap-Aktion](#) im Amazon EMR Management Guide.

Die folgenden Beispielskripts zeigen, wie eine Bootstrap-Aktionsdatei für Amazon EMR 6.x und 7.x erstellt wird:

Amazon EMR 7.x

my-bootstrap-action.sh Beispieldatei für Amazon EMR 7.x

YARN Um GPU Ressourcen mit Amazon EMR 7.x-Versionen zu verwalten, müssen Sie sie manuell CGroup v1 auf Ihrem Cluster bereitstellen. Sie können dies mit einem Bootstrap-Aktionsskript tun, wie in diesem Beispiel gezeigt.

```
#!/bin/bash
set -ex

sudo mkdir -p /spark-rapids-cgroup/devices
sudo mount -t cgroup -o devices cgroupv1-devices /spark-rapids-cgroup/devices
sudo chmod a+rwx -R /spark-rapids-cgroup
```

Amazon EMR 6.x

my-bootstrap-action.sh Beispieldatei für Amazon EMR 6.x

Für Amazon EMR 6.x-Versionen müssen Sie CGroup Berechtigungen für Ihren YARN Cluster öffnen. Sie können dies mit einem Bootstrap-Aktionsskript tun, wie in diesem Beispiel gezeigt.

```
#!/bin/bash
set -ex
```

```
sudo chmod a+rwx -R /sys/fs/cgroup/cpu,cpuacct
sudo chmod a+rwx -R /sys/fs/cgroup/devices
```

Ihren Cluster starten

Der letzte Schritt besteht darin, Ihren Cluster mit den oben genannten Cluster-Konfigurationen zu starten. Hier ist ein Beispielbefehl zum Starten eines Clusters von Amazon aus 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
```

Zugriff auf die Spark-Shell

Die Spark-Shell basiert auf dem Scala REPL (Read-Eval-Print-Loop). Damit können Sie Spark-Programme interaktiv erstellen und Arbeit an das Framework senden. Sie können auf die Spark-Shell zugreifen, indem Sie sich mit dem primären Knoten verbinden SSH und diesen aufrufen `spark-shell`. Weitere Informationen zum Herstellen einer Verbindung mit dem primären Knoten finden Sie unter [Verbindung zum primären Knoten herstellen SSH](#) im Amazon EMR Management Guide. Die folgenden Beispiele verwenden Apache HTTP Server-Zugriffsprotokolle, die in Amazon S3 gespeichert sind.

Note

Der Bucket in diesen Beispielen ist für Clients mit Zugriff auf USA Ost (Nord-Virginia) verfügbar.

Standardmäßig erstellt die Spark-Shell ihr eigenes [SparkContext](#) Objekt namens `sc`. Sie können diesen Kontext verwenden, wenn er innerhalb von erforderlich ist `REPL`. `sqlContext` ist auch in der Shell verfügbar und es ist ein [HiveContext](#).

Example Verwenden Sie die Spark-Shell, um die Vorkommen einer Zeichenkette in einer in Amazon S3 gespeicherten Datei zu zählen

Dieses Beispiel verwendet `sc` zum Lesen einer Textdatei in Amazon S3.

```
scala> sc
res0: org.apache.spark.SparkContext = org.apache.spark.SparkContext@404721db

scala> val textFile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/
impressions/dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark erstellt die `textFile` und die zugehörige [Datenstruktur](#). Im Beispiel wird dann die Anzahl der Zeilen in der Protokolldatei mit der Zeichenfolge "cartoonnetwork.com" gezählt:

```
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 Verwenden Sie die Python-basierte Spark-Shell, um die Vorkommen einer Zeichenkette in einer in Amazon S3 gespeicherten Datei zu zählen

Spark umfasst auch eine auf Python basierende Shell, `pyspark`, mit der Sie Prototypen von in Python geschriebenen Spark-Programmen entwickeln können. Genau wie `beispark-shell`, `pyspark` auf dem primären Knoten aufrufen; es hat auch dasselbe [SparkContext](#) Objekt.

```
>>> sc
<pyspark.context.SparkContext object at 0x7fe7e659fa50>
>>> textfile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/impressions/
dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark erstellt die `textFile` und die zugehörige [Datenstruktur](#). Im Beispiel wird dann die Anzahl der Zeilen in der Protokolldatei mit der Zeichenfolge "cartoonnetwork.com" gezählt.

```
>>> 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
```

Verwenden Sie Amazon SageMaker Spark für maschinelles Lernen

Wenn Sie Amazon EMR Version 5.11.0 und höher verwenden, wird die `aws-sagemaker-spark-sdk` Komponente zusammen mit Spark installiert. Diese Komponente installiert Amazon SageMaker Spark und die zugehörigen Abhängigkeiten für die Spark-Integration mit [Amazon SageMaker](#). Sie können Amazon SageMaker Spark verwenden, um Spark-Pipelines für maschinelles Lernen (ML) mithilfe von SageMaker Amazon-Stufen zu erstellen. Weitere Informationen finden Sie im [Amazon SageMaker Spark README](#) on GitHub and [Using Apache Spark with Amazon SageMaker](#) im Amazon SageMaker Developer Guide.

Eine Spark-Anwendung schreiben

[Spark](#)-Anwendungen können in Scala, Java und Python geschrieben werden. In der Dokumentation zu Apache Spark finden Sie im Thema [Spark-Beispiele](#) einige Beispiele für Spark-Anwendungen. Das Beispiel zum Schätzen von Pi, "Estimating Pi", ist unten in den drei nativ unterstützten Anwendungen dargestellt. Vollständige Beispiele finden Sie auch unter `$SPARK_HOME/examples` und unter [GitHub](#). Weitere Informationen darüber, wie man JARs für Spark baut, finden Sie im Thema [Schnellstart](#) in der Apache Spark-Dokumentation.

Scala

Um Scala-Kompatibilitätsprobleme zu vermeiden, empfehlen wir Ihnen, Spark-Abhängigkeiten für die richtige Scala-Version zu verwenden, wenn Sie eine Spark-Anwendung für einen EMR Amazon-Cluster kompilieren. Welche Scala-Version Sie verwenden sollten, hängt von der Version von Spark ab, die auf Ihrem Cluster installiert ist. Die EMR Amazon-Version 5.30.1 verwendet beispielsweise

Spark 2.4.5, das mit Scala 2.11 erstellt wurde. Wenn Ihr Cluster die EMR Amazon-Version 5.30.1 verwendet, verwenden Sie Spark-Abhängigkeiten für Scala 2.11. Weitere Informationen zu den von Spark verwendeten Scala-Versionen finden Sie in der [Apache-Spark-Dokumentation](#).

```
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)
```

Verbessern der Spark Leistung mit Amazon S3

Amazon EMR bietet Funktionen zur Leistungsoptimierung bei der Verwendung von Spark zum Abfragen, Lesen und Schreiben von in Amazon S3 gespeicherten Daten.

[S3 Select](#) kann die Abfrageleistung für CSV und JSON Dateien in einigen Anwendungen verbessern, indem die Verarbeitung auf Amazon S3 „heruntergeschoben“ wird.

Der EMRFS S3-optimierte Committer ist eine Alternative zur [OutputCommitter](#)-Klasse, die die Funktion für mehrteilige Uploads von verwendet, EMRFS um die Leistung beim Schreiben von Parquet-Dateien auf Amazon S3 mithilfe von SparkSQL, und Datasets zu verbessern. DataFrames

Themen

- [S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden](#)
- [Verwenden Sie den EMRFS S3-optimierten Committer](#)
- [Verwenden Sie das S3-optimierte Commit-Protokoll EMRFS](#)
- [Versuchen Sie Amazon S3 S3-Anfragen erneut mit EMRFS](#)

S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden

Mit EMR Amazon-Version 5.17.0 und höher können Sie [S3 Select](#) mit Spark auf Amazon verwenden. EMR S3 Select ermöglicht es Anwendungen, nur eine Teilmenge von Daten aus einem Objekt abzurufen. Für Amazon EMR wird die Rechenarbeit beim Filtern großer Datenmengen für die Verarbeitung vom Cluster auf Amazon S3 „heruntergeschoben“, wodurch die Leistung einiger

Anwendungen verbessert und die zwischen Amazon EMR und Amazon S3 übertragene Datenmenge reduziert werden kann.

S3 Select wird zusammen mit CSV JSON Dateien unterstützt, die `s3selectJSON` Werte `s3selectCSV` und zur Angabe des Datenformats verwenden. Weitere Informationen und Beispiele finden Sie unter [S3 Select in Ihrem Code angeben](#).

Ist S3 Select das Richtige für meine Anwendung?

Wir empfehlen, dass Sie Benchmark-Tests für Ihre Anwendungen im Vergleich mit und ohne S3 Select durchführen, um zu sehen, ob es für Ihre Anwendung geeignet sein könnte.

Verwenden Sie die folgenden Richtlinien, um zu bestimmen, ob Ihre Anwendung ein Kandidat für die Verwendung von S3 ist:

- Ihre Abfrage filtert mehr als die Hälfte des ursprünglichen Datensatzes.
- Ihre Netzwerkverbindung zwischen Amazon S3 und dem EMR Amazon-Cluster hat eine gute Übertragungsgeschwindigkeit und verfügbare Bandbreite. Amazon S3 komprimiert HTTP Antworten nicht, daher wird die Antwortgröße bei komprimierten Eingabedateien wahrscheinlich zunehmen.

Überlegungen und Einschränkungen

- Serverseitige Amazon S3 S3-Verschlüsselung mit vom Kunden bereitgestellten Verschlüsselungsschlüsseln (SSE-C) und clientseitige Verschlüsselung werden nicht unterstützt.
- Die Eigenschaft `AllowQuotedRecordDelimiters` wird nicht unterstützt. Wenn diese Eigenschaft angegeben ist, schlägt die Abfrage fehl.
- Es werden nur JSON Dateien im CSV -8-Format unterstütztUTF. Mehrzeilige CSVs werden nicht unterstützt.
- Es werden nur unkomprimierte oder gzip-Dateien unterstützt.
- Spark CSV und JSON Optionen `wiananValue`, und Optionen `positiveInfnegativeInf`, die sich auf beschädigte Datensätze beziehen (z. B. Failfast- und Dropmalformed-Modus), werden nicht unterstützt.
- Die Verwendung von Kommas als Tausender-Trennzeichen wird bei Dezimalzahlen nicht unterstützt. Beispielweise ist `10,000` ungültig, `10000` wird unterstützt.
- Kommentarzeichen auf der letzten Zeile werden nicht unterstützt.
- Leere Zeilen am Ende einer Datei werden nicht verarbeitet.

- Die folgenden Filter werden nicht an Amazon S3 ausgelagert:
 - Aggregatfunktionen wie COUNT() und SUM().
 - Filter, die ein Attribut über einen CAST()-Aufruf konvertieren. Beispiel, CAST(stringColumn as INT) = 1.
 - Filter mit einem Attribut, das ein Objekt ist oder einem zusammengesetzten Attribut. Beispiel, intArray[1] = 1, objectColumn.objectNumber = 1.
 - Filter, deren Wert nicht literal ist. Beispiel: intColumn1 = intColumn2
 - Es werden nur die [von S3 Select unterstützten Datentypen](#) unterstützt, und es gelten die dokumentierten Einschränkungen.

S3 Select in Ihrem Code angeben

Die folgenden Beispiele zeigen, wie S3 Select für die CSV Verwendung von ScalaSQL, R und PySpark angegeben wird. Sie können S3 Select for JSON auf die gleiche Weise verwenden. Eine Liste der Optionen, der Standardwerte und der Einschränkungen finden Sie unter [Optionen](#).

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")
```

Optionen

Die folgenden Optionen sind verfügbar, wenn `s3selectCSV` und `s3selectJSON` verwendet wird. Wenn nichts angegeben wird, werden die Standardwerte verwendet.

Optionen mit S3Select CSV

| Option | Standard | Verwendung |
|--------------------------|----------------------|--|
| <code>compression</code> | <code>"none"</code> | Gibt an, ob Komprimierung verwendet wird. <code>"gzip"</code> wird neben <code>"none"</code> als einzige Einstellung unterstützt. |
| <code>delimiter</code> | <code>","</code> | Gibt den Feldbegrenzer an. |
| <code>quote</code> | <code>'\"'</code> | Gibt das Zeichen an, das als Anführungszeichen verwendet wird. Die Angabe einer leeren Zeichenfolge wird nicht unterstützt und führt zu einem XML fehlerhaften Fehler. |
| <code>escape</code> | <code>'\\'</code> | Gibt das Zeichen an, das als Escape-Zeichen verwendet wird. |
| <code>header</code> | <code>"false"</code> | <code>"false"</code> gibt an, dass es keine Kopfzeile gibt. <code>"true"</code> gibt an, dass die erste Zeile die Kopfzeile ist. Es werden ausschließlich Kopfzeilen |

| Option | Standard | Verwendung |
|------------------------|----------|--|
| | | in der ersten Zeile unterstützt, und leere Zeilen vor der Kopfzeile werden nicht unterstützt. |
| Kommentar | "#" | Gibt das Zeichen an, das als Kommentarzeichen verwendet wird. Der Kommentarindikator kann nicht deaktiviert werden. Dies bedeutet, dass der Wert <code>\u0000</code> nicht unterstützt wird. |
| <code>nullValue</code> | "" | |

Optionen mit S3Select JSON

| Option | Standard | Verwendung |
|--------------------------|----------|--|
| <code>compression</code> | "none" | Gibt an, ob Komprimierung verwendet wird. "gzip" wird neben "none" als einzige Einstellung unterstützt. |
| <code>multiline</code> | „false“ | "false" gibt an, dass JSON das das LINES Format S3 Select hat, was bedeutet, dass jede Zeile in den Eingabedaten ein einzelnes JSON Objekt enthält. "true" gibt an, dass JSON es sich um das DOCUMENT S3-Select-Format handelt, was bedeutet, dass sich ein JSON Objekt in den Eingabedaten über mehrere Zeilen erstrecken kann. |

Verwenden Sie den EMRFS S3-optimierten Committer

Der EMRFS S3-optimierte Committer ist eine alternative [OutputCommitter](#) Implementierung, die für das Schreiben von Dateien in Amazon S3 bei Verwendung optimiert ist. EMRFS Der EMRFS S3-optimierte Committer verbessert die Anwendungsleistung, indem er Listen- und Umbenennungsvorgänge vermeidet, die in Amazon S3 während der Job- und Task-Commit-Phasen ausgeführt werden. Der Committer ist mit EMR Amazon-Version 5.19.0 und höher verfügbar und ab Amazon EMR 5.20.0 standardmäßig aktiviert. Der Committer wird für Spark-Jobs verwendet, die Spark oder Datasets verwenden. SQL DataFrames Ab Amazon EMR 6.4.0 kann dieser Committer für alle gängigen Formate verwendet werden ORC, einschließlich Parquet und textbasierte Formate (einschließlich CSV und). JSON Für Versionen vor Amazon EMR 6.4.0 wird nur das Parquet-Format unterstützt. Es gibt Situationen, in denen der Committer nicht verwendet wird. Weitere Informationen finden Sie unter [Anforderungen für den S3-optimierten EMRFS Committer](#).

Themen

- [Anforderungen für den S3-optimierten EMRFS Committer](#)
- [Der EMRFS S3-optimierte Committer und mehrteilige Uploads](#)
- [Überlegungen zur Auftragsoptimierung](#)
- [Aktivieren Sie den EMRFS S3-optimierten Committer für Amazon 5.19.0 EMR](#)


Anforderungen für den S3-optimierten EMRFS Committer

Der EMRFS S3-optimierte Committer wird verwendet, wenn die folgenden Bedingungen erfüllt sind:

- Sie führen Spark-Jobs aus, die Spark SQL oder Datasets verwenden DataFrames, um Dateien in Amazon S3 zu schreiben. Ab Amazon EMR 6.4.0 kann dieser Committer für alle gängigen Formate verwendet werden ORC, einschließlich Parquet und textbasierte Formate (einschließlich CSV und). JSON Für Versionen vor Amazon EMR 6.4.0 wird nur das Parquet-Format unterstützt.
- Mehrteilige Uploads sind in Amazon aktiviert. EMR Dies ist die Standardeinstellung. Weitere Informationen finden Sie unter [Der EMRFS S3-optimierte Committer und mehrteilige Uploads](#).
- Die integrierte Dateiformatunterstützung von Spark wird verwendet. Die integrierte Dateiformatunterstützung wird unter folgenden Umständen verwendet:
 - Bei Hive-Metastore-Tabellen `spark.sql.hive.convertMetastoreParquet` ist wann `true` für Parquet-Tabellen oder `spark.sql.hive.convertMetastoreOrc` `true` für Orc-Tabellen mit Amazon EMR 6.4.0 oder höher auf eingestellt. Dies sind die Standardeinstellungen.

- Wenn Aufträge in Datenquellen oder Tabellen im Dateiformat schreiben, wird beispielsweise die Zieltabelle mit der `USING parquet` Klausel erstellt.
- Wenn Aufträge in nicht partitionierte Hive-Metastore-Parquet-Tabellen schreiben. Eine bekannte Einschränkung besteht darin, dass die in Spark integrierte Parquet-Unterstützung keine partitionierten Hive-Tabellen unterstützt. Weitere Informationen finden Sie unter [Konvertierung von Hive Metastore Parquet-Tabellen](#) im Apache Spark and Datasets Guide. SQL DataFrames
- Spark-Job-Operationen, die beispielsweise `${table_location}/k1=v1/k2=v2/` in einen Standardspeicherort für Partitionen schreiben, verwenden den Committer. Der Committer wird nicht verwendet, wenn ein Job-Vorgang in einen benutzerdefinierten Partitionsspeicherort schreibt, z. B. wenn mit dem `ALTER TABLE SQL`-Befehl ein benutzerdefinierter Partitionsspeicherort festgelegt wurde.
- Für Spark müssen die folgenden Werte verwendet werden:
 - Der Eigenschaft `spark.sql.parquet.fs.optimized.committer.optimization-enabled` muss auf `true` eingestellt sein. Dies ist die Standardeinstellung bei Amazon EMR 5.20.0 und höher. Bei Amazon EMR 5.19.0 ist der Standardwert `false`. Weitere Informationen zum Konfigurieren dieses Werts finden Sie unter [Aktivieren Sie den EMRFS S3-optimierten Committer für Amazon 5.19.0 EMR](#).
 - Beim Schreiben in nicht partitionierte Hive-Metastore-Tabellen werden nur die Dateiformate Parquet und Orc unterstützt. `spark.sql.hive.convertMetastoreParquet` muss auf `true` gesetzt sein, wenn in nicht partitionierte Parquet Hive-Metastore-Tabellen geschrieben werden. `spark.sql.hive.convertMetastoreOrc` muss auf `true` gesetzt sein, wenn in nicht partitionierte Orc Hive Metastore-Tabellen geschrieben werden. Dies sind die Standardeinstellungen.
 - `spark.sql.parquet.output.committer.class` muss auf `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter` festgelegt sein. Dies ist die Standardeinstellung.
 - `spark.sql.sources.commitProtocolClass` muss auf oder gesetzt sein.
 - `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`
 - `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`
 - `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`
 - die Standardeinstellung für die Amazon EMR 5.x-Serie Version 5.30.0 und höher und für die Amazon EMR 6.x-Serie Version 6.2.0 und höher.
 - `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`
 - die Standardeinstellung für frühere EMR Amazon-Versionen.

- Wenn Spark-Aufträge partitionierte Parquet-Datasets durch dynamische Partitionsspalten überschreiben, dann müssen die Schreiboption `partitionOverwriteMode` und `spark.sql.sources.partitionOverwriteMode` auf `static` eingestellt sein. Dies ist die Standardeinstellung.

 Note

Die Schreiboption `partitionOverwriteMode` wurde mit Spark 2.4.0 eingeführt. Für Spark-Version 2.3.2, die in der EMR Amazon-Version 5.19.0 enthalten ist, legen Sie die Eigenschaft fest. `spark.sql.sources.partitionOverwriteMode`

Fälle, in denen der EMRFS S3-optimierte Committer nicht verwendet wird

Im Allgemeinen wird der EMRFS S3-optimierte Committer in den folgenden Situationen nicht verwendet.

| Situation | Warum der Committer nicht verwendet wird |
|--|--|
| Wenn Sie schreiben an HDFS | Der Committer unterstützt nur das Schreiben in Amazon S3 mithilfe von EMRFS. |
| Bei Verwendung des S3A-Dateisystems | Der Committer unterstützt nur EMRFS |
| Wenn Sie MapReduce oder Spark verwenden
RDD API | Der Committer unterstützt nur die Verwendung von SparkSQL, DataFrame, oder Dataset APIs. |

Die folgenden Scala-Beispiele zeigen einige zusätzliche Situationen, die verhindern, dass der EMRFS S3-optimierte Committer vollständig (das erste Beispiel) und teilweise (das zweite Beispiel) verwendet wird.

Example – Dynamischer Partitionsüberschreibmodus

Das folgende Scala-Beispiel weist Spark an, einen anderen Commit-Algorithmus zu verwenden, wodurch die Verwendung des S3-optimierten Committers vollständig verhindert wird. EMRFS Der Code legt die Eigenschaft `partitionOverwriteMode` auf `dynamic` fest, sodass nur die Partitionen überschrieben werden, auf die Sie Daten schreiben. Anschließend werden dynamische

Partitionsspalten durch `partitionBy` angegeben und der Schreibmodus ist auf `overwrite` eingestellt.

```
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")
```

Sie müssen alle drei Einstellungen konfigurieren, um die Verwendung des S3-optimierten Committers zu vermeiden. EMRFS Wenn Sie dies tun, führt Spark einen anderen Commit-Algorithmus aus, der im Commit-Protokoll von Spark angegeben ist. Für Amazon EMR 5.x-Versionen vor 5.30.0 und für Amazon EMR 6.x-Versionen vor 6.2.0 verwendet das Commit-Protokoll das Staging-Verzeichnis von Spark, ein temporäres Verzeichnis, das unter dem Ausgabeverzeichnis erstellt wird, das mit `spark-staging` beginnt. Der Algorithmus benennt Partitionsverzeichnisse nacheinander um, was sich negativ auf die Leistung auswirken kann. Weitere Informationen zu den EMR Amazon-Versionen 5.30.0 und höher sowie 6.2.0 und höher finden Sie unter [Verwenden Sie das S3-optimierte Commit-Protokoll EMRFS](#)

Der Algorithmus in Spark 2.4.0 folgt diesen Schritten:

1. Aufgabenversuche schreiben ihre Ausgabe in Partitionsverzeichnisse unterhalb des Staging-Verzeichnisses von Spark, beispielsweise `${outputLocation}/spark-staging-${jobID}/k1=v1/k2=v2/`.
2. Für jede geschriebene Partition verfolgt der Aufgabenversuch die relativen Partitionspfade, z. B. `k1=v1/k2=v2`.
3. Wenn eine Aufgabe erfolgreich abgeschlossen wurde, stellt sie dem Treiber alle zugehörigen Partitionspfade bereit, die von ihr nachverfolgt wurden.
4. Nachdem alle Aufgaben abgeschlossen wurden, sammelt die Auftrags-Commit-Phase alle die Partitionsverzeichnisse, die von erfolgreichen Aufgabenversuchen unter dem Staging-Verzeichnis von Spark geschrieben wurden. Spark benennt jedes dieser Verzeichnisse mithilfe von Verzeichnisstruktur-Umbenennungsoperationen sequenziell bis zu ihrem endgültigen Ausgabespeicherort um.
5. Das Staging-Verzeichnis wird gelöscht, bevor der Auftrags-Commit-Phase abgeschlossen ist.

Example – Benutzerdefinierter Partitionsspeicherort

In diesem Beispiel wird der Scala-Code in zwei Partitionen eingefügt. Eine Partition verfügt über einen benutzerdefinierten Partitionsspeicherort. Die andere Partition verwendet den Standard-Partitionsspeicherort. Der EMRFS S3-optimierte Committer wird nur zum Schreiben der Aufgabenausgabe auf die Partition verwendet, die den Standardspeicherort der Partition verwendet.

```
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)
```

Der Scala-Code erstellt die folgenden Amazon-S3-Objekte:

```
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_$folder$
```

Beim Schreiben in Partitionen an benutzerdefinierten Speicherorten verwendet Spark einen Commit-Algorithmus, ähnlich wie im vorherigen Beispiel. Dies wird im Folgenden beschrieben. Genau wie bei dem vorherigen Beispiel führt der Algorithmus zu sequenziellen Umbenennungen, wodurch die Leistung beeinträchtigt werden kann.

1. Beim Schreiben von Ausgabe in eine Partition an einem benutzerdefinierten Speicherort werden Aufgaben in eine Datei unter dem Staging-Verzeichnis von Spark geschrieben, das unter dem endgültigen Ausgabespeicherort erstellt wird. Der Name der Datei enthält zum Schutz vor Dateikollisionen einen UUID zufälligen Wert. Der Aufgabe-Versuch verfolgt jede Datei zusammen mit dem gewünschten endgültigen Ausgabepfad nach.
2. Wenn eine Aufgabe erfolgreich abgeschlossen wird, stellt sie dem Treiber die Dateien und die für sie gewünschten endgültigen Ausgabepfade bereit.
3. Nachdem alle Aufgaben beendet wurden, benennt die Auftrags-Commit-Phase alle Dateien, die für Partitionen an benutzerdefinierten Speicherorten geschrieben wurden, sequentiell in ihre endgültigen Ausgabepfade um.
4. Das Staging-Verzeichnis wird gelöscht, bevor der Auftrags-Commit-Phase abgeschlossen ist.

Der EMRFS S3-optimierte Committer und mehrteilige Uploads

Um den EMRFS S3-optimierten Committer verwenden zu können, müssen Sie mehrteilige Uploads für Amazon aktivieren. EMR Mehrteilige Uploads sind standardmäßig aktiviert. Sie können diese Option bei Bedarf erneut aktivieren. Weitere Informationen finden [Sie unter Mehrteiligen Upload für Amazon S3 konfigurieren](#) im Amazon EMR Management Guide.

Der EMRFS S3-optimierte Committer nutzt die transaktionsähnlichen Eigenschaften mehrteiliger Uploads, um sicherzustellen, dass Dateien, die durch Versuche von Aufgaben geschrieben wurden, beim Commit der Aufgabe nur am Ausgabespeicherort des Jobs erscheinen. Durch die Verwendung von mehrteiligen Uploads auf diese Weise verbessert der Committer die Leistung des Task-Commits im Vergleich zur Standardalgorithmusversion 2. FileOutputCommitter Bei der Verwendung des EMRFS S3-optimierten Committers sind einige wichtige Unterschiede zum herkömmlichen Verhalten bei mehrteiligen Uploads zu berücksichtigen:

- Mehrteilige Uploads werden immer ausgeführt, unabhängig von der Dateigröße. Dies unterscheidet sich vom Standardverhalten von EMRFS, bei dem die `fs.s3n.multipart.uploads.split.size` Eigenschaft die Dateigröße steuert, bei der mehrteilige Uploads ausgelöst werden.

- Mehrteilige Uploads verbleiben für einen längeren Zeitraum in einem Status, in dem sie nicht abgeschlossen sind, bis die Aufgabe übertragen oder abgebrochen wird. Dies unterscheidet sich vom Standardverhalten, EMRFS bei dem ein mehrteiliger Upload abgeschlossen wird, wenn eine Aufgabe das Schreiben einer bestimmten Datei abgeschlossen hat.

Aufgrund dieser Unterschiede ist es wahrscheinlicher, dass unvollständige mehrteilige Uploads zurückbleiben, wenn ein Spark Executor JVM abstürzt oder beendet wird, während Aufgaben ausgeführt werden und Daten auf Amazon S3 geschrieben werden. Aus diesem Grund sollten Sie bei der Verwendung des EMRFS S3-optimierten Committers unbedingt die Best Practices für den Umgang mit fehlgeschlagenen mehrteiligen Uploads befolgen. Weitere Informationen finden Sie unter [Bewährte Methoden](#) für die Arbeit mit Amazon S3 S3-Buckets im Amazon EMR Management Guide.

Überlegungen zur Auftragsoptimierung

Der EMRFS S3-optimierte Committer verbraucht für jede Datei, die bei einem Task-Versuch geschrieben wurde, eine geringe Menge an Speicher, bis die Aufgabe festgeschrieben oder abgebrochen wird. Bei den meisten Aufträgen ist die Menge des belegten Speichers vernachlässigbar. Bei Aufträgen mit Aufgaben mit langer Ausführungsdauer, die eine große Anzahl von Dateien schreiben, macht sich der Arbeitsspeicher, den der Committer benötigt, bemerkbar, und dies erfordert möglicherweise Anpassungen an den für Spark Executor zugewiesenen Arbeitsspeicher. Sie können die Eigenschaft `spark.executor.memory` verwenden, um den Executor-Arbeitsspeicher anzupassen. Als Faustregel gilt: für jeweils 100.000 Dateien, die eine einzelne Aufgabe schreiben, werden in der Regel zusätzlich 100 MB Arbeitsspeicher benötigt. Weitere Informationen finden Sie unter [Anwendungseigenschaften](#) in der Dokumentation zur Konfiguration von Apache Spark.

Aktivieren Sie den EMRFS S3-optimierten Committer für Amazon 5.19.0 EMR

Wenn Sie Amazon EMR 5.19.0 verwenden, können Sie die `spark.sql.parquet.fs.optimized.committer.optimization-enabled` Eigenschaft manuell auf `true` festlegen, wenn Sie einen Cluster erstellen, oder von Spark aus, wenn Sie Amazon verwenden. EMR

Aktivierung des EMRFS S3-optimierten Committers bei der Erstellung eines Clusters

Verwenden Sie die Konfigurationsklassifizierung `spark-defaults`, um die `spark.sql.parquet.fs.optimized.committer.optimization-enabled`-Eigenschaften auf `true` festzulegen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktivierung des EMRFS S3-optimierten Committers von Spark

Sie können `spark.sql.parquet.fs.optimized.committer.optimization-enabled` auf `true` festlegen, indem Sie dies in einer SparkConf fest kodieren, als `--conf`-Parameter in der Spark-Shell übergeben, die Tools `spark-submit` bzw. `spark-sql` verwenden oder dies in `conf/spark-defaults.conf` angeben. Weiter Informationen finden Sie unter [Spark-Konfiguration](#) in der Apache-Spark-Dokumentation.

Im folgenden Beispiel wird gezeigt, wie Sie den Committer während der Ausführung eines `spark-sql`-Befehls aktivieren.

```
spark-sql \  
  --conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \  
  -e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

Verwenden Sie das S3-optimierte Commit-Protokoll EMRFS

Das EMRFS S3-optimierte Commit-Protokoll ist eine alternative [FileCommitProtocol](#) Implementierung, die für das Schreiben von Dateien mit dynamischem Spark-Partitionsüberschreiben in Amazon S3 optimiert ist, wenn sie verwendet werden. EMRFS Das Protokoll verbessert die Anwendungsleistung, indem Umbenennungsvorgänge in Amazon S3 während der Commit-Phase für das dynamische Überschreiben von Spark-Partitionen vermieden werden.

Beachten Sie, dass [Verwenden Sie den EMRFS S3-optimierten Committer](#) auch die Leistung verbessert, indem es Umbenennungsvorgänge vermeidet. Es funktioniert jedoch nicht für Fälle dynamischer Partitionsüberschreibungen, während die Verbesserungen des Commit-Protokolls nur auf Fälle dynamischer Partitionsüberschreibungen abzielen.

Das Commit-Protokoll ist mit EMR Amazon-Version 5.30.0 und höher und 6.2.0 und höher verfügbar und standardmäßig aktiviert. Amazon EMR hat ab Version 5.31.0 eine Verbesserung der Parallelität hinzugefügt. Das Protokoll wird für Spark-Jobs verwendet, die Spark oder Datasets SQL verwenden. DataFrames Es gibt Situationen, in denen das Commit Protokoll nicht verwendet wird. Weitere Informationen finden Sie unter [Anforderungen für das EMRFS S3-optimierte Commit-Protokoll](#).

Themen

- [Anforderungen für das EMRFS S3-optimierte Commit-Protokoll](#)
- [Das EMRFS S3-optimierte Commit-Protokoll und mehrteilige Uploads](#)
- [Überlegungen zur Auftragsoptimierung](#)

Anforderungen für das EMRFS S3-optimierte Commit-Protokoll

Das EMRFS S3-optimierte Commit-Protokoll wird verwendet, wenn die folgenden Bedingungen erfüllt sind:

- Sie führen Spark-Jobs aus, die Spark oder Datasets verwenden SQL DataFrames, um partitionierte Tabellen zu überschreiben.
- Sie führen Spark-Jobs aus, deren Partitionsüberschreibmodus `dynamic` ist.
- Mehrteilige Uploads sind in Amazon aktiviert. EMR Dies ist die Standardeinstellung. Weitere Informationen finden Sie unter [Das EMRFS S3-optimierte Commit-Protokoll und mehrteilige Uploads](#).
- Der Dateisystem-Cache für EMRFS ist aktiviert. Dies ist die Standardeinstellung. Vergewissern Sie sich, dass die Einstellung `fs.s3.impl.disable.cache` auf `false` gesetzt ist.
- Die integrierte Datenquellenunterstützung von Spark wird verwendet. Die integrierte Datenquellenunterstützung wird in den folgenden Fällen verwendet:
 - Wenn Aufträge in integrierte Datenquellen oder Tabellen schreiben.
 - Wenn Aufträge in die Parquet-Tabelle des Hive-Metastores schreiben. Das passiert, wenn `spark.sql.hive.convertInsertingPartitionedTable` und `spark.sql.hive.convertMetastoreParquet` beide auf „wahr“ gesetzt sind. Dies sind die Standardeinstellungen.
 - Wenn Jobs in die ORC Hive-Metastore-Tabelle schreiben. Das passiert, wenn `spark.sql.hive.convertInsertingPartitionedTable` und `spark.sql.hive.convertMetastoreOrc` beide auf `true` gesetzt sind. Dies sind die Standardeinstellungen.
- Spark-Auftrags-Vorgänge, die z. B. in einen Standard-Partitionsspeicherort schreiben, z. B. `${table_location}/k1=v1/k2=v2/`, verwenden das Commit-Protokoll. Das Protokoll wird nicht verwendet, wenn ein Jobvorgang in einen benutzerdefinierten Partitionsspeicherort schreibt, beispielsweise wenn ein benutzerdefinierter Partitionsspeicherort mit dem Befehl `ALTER TABLE SQL` festgelegt wird.
- Für Spark müssen die folgenden Werte verwendet werden:
 - muss `spark.sql.sources.commitProtocolClass` auf `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` festgelegt sein. Dies ist die Standardeinstellung für EMR Amazon-Versionen 5.30.0 und höher sowie 6.2.0 und höher.

- Die Schreiboption `partitionOverwriteMode` oder `spark.sql.sources.partitionOverwriteMode` muss auf `dynamic` gesetzt sein. Die Standardeinstellung lautet `static`.

Note

Die Schreiboption `partitionOverwriteMode` wurde mit Spark 2.4.0 eingeführt. Für Spark-Version 2.3.2, die in der EMR Amazon-Version 5.19.0 enthalten ist, legen Sie die Eigenschaft fest. `spark.sql.sources.partitionOverwriteMode`

- Wenn Spark-Aufträge in die Hive-Metastore-Parquet-Tabelle überschreiben muss `spark.sql.hive.convertMetastoreParquet`, `spark.sql.hive.convertInsertingPartitionedTable` und `spark.sql.hive.convertMetastore.partitionOverwriteMode` auf `true` gesetzt werden. Dies sind die Standardeinstellungen.
- Wenn Spark-Jobs die ORC Hive-Metastore-Tabelle überschreiben, `spark.sql.hive.convertMetastoreOrc` `spark.sql.hive.convertInsertingPartitionedTable` und muss auf `true` gesetzt werden. `spark.sql.hive.convertMetastore.partitionOverwriteMode true` Dies sind die Standardeinstellungen.

Example – Dynamischer Partitionsüberschreibmodus

In diesem Scala-Beispiel wird die Optimierung ausgelöst. Als erstes setzen Sie die `partitionOverwriteMode`-Eigenschaft auf `dynamic`. Dadurch werden nur die Partitionen überschrieben, auf die Sie Daten schreiben. Anschließend geben Sie dynamische Partitionsspalten mit `partitionBy` an und legen den Schreibmodus auf `overwrite` fest.

```
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://"
```


Wenn das EMRFS S3-optimierte Commit-Protokoll nicht verwendet wird

Im Allgemeinen funktioniert das EMRFS S3-optimierte Commit-Protokoll genauso wie das SQL Open-Source-Standard-Spark-Commit-Protokoll.

`org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`
In den folgenden Situationen wird keine Optimierung durchgeführt.

| Situation | Warum das Commit-Protokoll nicht verwendet wird |
|--|---|
| Wenn Sie schreiben an HDFS | Das Commit-Protokoll unterstützt nur das Schreiben in Amazon S3 mithilfe von EMRFS. |
| Bei Verwendung des S3A-Dateisystems | Das Commit-Protokoll unterstützt nur EMRFS. |
| Wenn Sie MapReduce oder Sparks verwenden
RDD API | Das Commit-Protokoll unterstützt nur die Verwendung von Spark SQL DataFrame,, oder Dataset APIs. |
| Wenn das dynamische Überschreiben der Partition nicht ausgelöst wird | Das Commit-Protokoll optimiert nur Fälle, in denen dynamische Partitionen überschrieben werden. Für andere Fälle siehe Verwenden Sie den EMRFS S3-optimierten Committer . |

Die folgenden Scala-Beispiele veranschaulichen einige zusätzliche Situationen, in die das EMRFS S3-optimierte Commit-Protokoll delegiert. `SQLHadoopMapReduceCommitProtocol`

Example – Dynamischer Partitionsüberschreibmodus mit benutzerdefiniertem Partitionsspeicherort

In diesem Beispiel überschreiben die Scala-Programme zwei Partitionen im dynamischen Partitionsüberschreibmodus. Eine Partition verfügt über einen benutzerdefinierten Partitionsspeicherort. Die andere Partition verwendet den Standard-Partitionsspeicherort. Das EMRFS S3-optimierte Commit-Protokoll verbessert nur die Partition, die den Standardspeicherort der Partition verwendet.

```
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")

```

Der Scala-Code erstellt die folgenden Amazon-S3-Objekte:

```

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

Das Schreiben in benutzerdefinierte Partitionsspeicherorte in früheren Spark-Versionen kann zu Datenverlust führen. In diesem Beispiel `dt= '2019-01-28'` würde die Partition verloren

gehen. [Weitere Informationen finden Sie unter SPARK -35106](#). Dies wurde in EMR Amazon-Version 5.33.0 und höher behoben, mit Ausnahme von 6.0.x und 6.1.x.

Beim Schreiben in Partitionen an benutzerdefinierten Speicherorten verwendet Spark einen Commit-Algorithmus, ähnlich wie im vorherigen Beispiel. Dies wird im Folgenden beschrieben. Genau wie bei dem vorherigen Beispiel führt der Algorithmus zu sequenziellen Umbenennungen, wodurch die Leistung beeinträchtigt werden kann.

Der Algorithmus in Spark 2.4.0 folgt diesen Schritten:

1. Beim Schreiben von Ausgabe in eine Partition an einem benutzerdefinierten Speicherort werden Aufgaben in eine Datei unter dem Staging-Verzeichnis von Spark geschrieben, das unter dem endgültigen Ausgabespeicherort erstellt wird. Der Name der Datei enthält zum Schutz vor Dateikollisionen einen zufälligen WertUUID. Der Aufgabe-Versuch verfolgt jede Datei zusammen mit dem gewünschten endgültigen Ausgabepfad nach.
2. Wenn eine Aufgabe erfolgreich abgeschlossen wird, stellt sie dem Treiber die Dateien und die für sie gewünschten endgültigen Ausgabepfade bereit.
3. Nachdem alle Aufgaben beendet wurden, benennt die Auftrags-Commit-Phase alle Dateien, die für Partitionen an benutzerdefinierten Speicherorten geschrieben wurden, sequentiell in ihre endgültigen Ausgabepfade um.
4. Das Staging-Verzeichnis wird gelöscht, bevor der Auftrags-Commit-Phase abgeschlossen ist.

Das EMRFS S3-optimierte Commit-Protokoll und mehrteilige Uploads

Um die Optimierung für dynamisches Überschreiben von Partitionen im EMRFS S3-optimierten Commit-Protokoll nutzen zu können, müssen mehrteilige Uploads in Amazon aktiviert sein. EMR Mehrteilige Uploads sind standardmäßig aktiviert. Sie können diese Option bei Bedarf erneut aktivieren. Weitere Informationen finden [Sie unter Mehrteiligen Upload für Amazon S3 konfigurieren](#) im Amazon EMR Management Guide.

Beim dynamischen Überschreiben von Partitionen verwendet das EMRFS S3-optimierte Commit-Protokoll die transaktionsähnlichen Eigenschaften von mehrteiligen Uploads, um sicherzustellen, dass Dateien, die durch Taskversuche geschrieben wurden, beim Commit des Jobs nur am Ausgabespeicherort des Jobs erscheinen. Durch die Verwendung von mehrteiligen Uploads auf diese Weise verbessert das Commit-Protokoll die Leistung von Auftrags-Commits im Vergleich zum Standard-SQLHadoopMapReduceCommitProtocol. Bei der Verwendung des EMRFS S3-

optimierten Commit-Protokolls müssen einige wichtige Unterschiede zum herkömmlichen Verhalten bei mehrteiligen Uploads berücksichtigt werden:

- Mehrteilige Uploads werden immer ausgeführt, unabhängig von der Dateigröße. Dies unterscheidet sich vom Standardverhalten von EMRFS, bei dem die `fs.s3n.multipart.uploads.split.size` Eigenschaft die Dateigröße steuert, bei der mehrteilige Uploads ausgelöst werden.
- Mehrteilige Uploads verbleiben für einen längeren Zeitraum in einem Status, in dem sie nicht abgeschlossen sind, bis die Aufgabe übertragen oder abgebrochen wird. Dies unterscheidet sich vom Standardverhalten, EMRFS bei dem ein mehrteiliger Upload abgeschlossen wird, wenn eine Aufgabe das Schreiben einer bestimmten Datei abgeschlossen hat.

Aufgrund dieser Unterschiede ist es wahrscheinlicher, dass unvollständige mehrteilige Uploads zurückbleiben, wenn ein Spark Executor JVM JVM abstürzt oder beendet wird, während Aufgaben ausgeführt werden und Daten in Amazon S3 geschrieben werden, oder wenn ein Spark-Treiber abstürzt oder beendet wird, während ein Job ausgeführt wird. Aus diesem Grund sollten Sie bei der Verwendung des EMRFS S3-optimierten Commit-Protokolls unbedingt die Best Practices für den Umgang mit fehlgeschlagenen mehrteiligen Uploads befolgen. Weitere Informationen finden Sie unter [Bewährte Methoden](#) für die Arbeit mit Amazon S3 S3-Buckets im Amazon EMR Management Guide.

Überlegungen zur Auftragsoptimierung

Auf Spark-Executoren verbraucht das EMRFS S3-optimierte Commit-Protokoll eine geringe Menge an Speicher für jede Datei, die bei einem Task-Versuch geschrieben wurde, bis die Aufgabe festgeschrieben oder abgebrochen wird. Bei den meisten Aufträgen ist die Menge des belegten Speichers vernachlässigbar.

Auf Spark-Treibern benötigt das EMRFS S3-optimierte Commit-Protokoll Speicher, um die Metadateninformationen jeder übergebenen Datei zu speichern, bis der Job festgeschrieben oder abgebrochen wird. Bei den meisten Aufträgen ist die standardmäßige Speichereinstellung des Spark-Treibers vernachlässigbar.

Bei Aufträgen mit Aufgaben mit langer Laufzeit, die eine große Anzahl von Dateien schreiben, kann der Speicherverbrauch des Commit-Protokolls spürbar sein und Anpassungen des für Spark zugewiesenen Speichers erfordern, insbesondere für Spark-Ausführer. Sie können den Speicher mithilfe der `spark.driver.memory`-Eigenschaft für Spark-Treiber und der Eigenschaft für `spark.executor.memory`-Spark-Aufträge optimieren. Als Faustregel gilt: für jeweils 100.000 Dateien, die eine einzelne Aufgabe schreiben, werden in der Regel zusätzlich 100 MB Arbeitsspeicher

benötigt. Weiter Informationen finden Sie unter [Anwendungseigenschaften](#) in der Dokumentation zur Konfiguration von Apache Spark.

Versuchen Sie Amazon S3 S3-Anfragen erneut mit EMRFS

Dieses Thema enthält Informationen zu den Wiederholungsstrategien, die Sie verwenden können, wenn Sie Anfragen an Amazon S3 stellen EMRFS. Wenn Ihre Anforderungsrate steigt, versucht S3, zu skalieren, um die neue Rate zu unterstützen. Während dieses Vorgangs kann S3 Anforderungen drosseln und einen 503 Slow Down-Fehler zurückgeben. Um die Erfolgsquote Ihrer S3-Anforderungen zu verbessern, können Sie Ihre Wiederholungsstrategie anpassen, indem Sie Eigenschaften in Ihrer `emrfs-site`-Konfiguration konfigurieren.

Sie können Ihre Wiederholungsstrategie wie folgt anpassen.

- Erhöhen Sie das maximale Wiederholungslimit für die standardmäßige exponentielle Backoff-Wiederholungsstrategie.
- Aktivieren und konfigurieren Sie die Wiederholungsstrategie additiv erhöhen/multiplikativ verringern (). AIMD AIMD wird für EMR Amazon-Versionen 6.4.0 und höher unterstützt.

Verwenden Sie die standardmäßige exponentielle Backoff-Strategie

EMRFS verwendet standardmäßig eine exponentielle Backoff-Strategie, um Amazon S3 S3-Anfragen erneut zu versuchen. Das Standardlimit für EMRFS Wiederholungsversuche ist 15. Um einen 503 Slow Down-S3-Fehler zu vermeiden, können Sie das Wiederholungslimit erhöhen, wenn Sie einen neuen Cluster, in einem laufenden Cluster oder zur Laufzeit der Anwendung erstellen.

Um das Wiederholungslimit zu erhöhen, müssen Sie den Wert für `fs.s3.maxRetries` in Ihrer `emrfs-site`-Konfiguration ändern. Die folgende Beispielkonfiguration setzt `fs.s3.maxRetries` auf einen benutzerdefinierten Wert von 30.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.maxRetries": "30"
    }
  }
]
```

Weitere Informationen zum Arbeiten mit Konfigurationsobjekten finden Sie unter [Anwendungen konfigurieren](#).

Verwenden Sie die AIMD Wiederholungsstrategie

EMRFS unterstützt mit EMR Amazon-Version 6.4.0 und höher eine alternative Wiederholungsstrategie, die auf einem additiven Zunahme/Multiplikative-Verringerungs-Modell () basiert. AIMD Die AIMD Wiederholungsstrategie ist besonders nützlich, wenn Sie mit großen EMR Amazon-Clustern arbeiten.

AIMD berechnet anhand von Daten über kürzlich erfolgreiche Anfragen eine benutzerdefinierte Anforderungsrate. Diese Strategie verringert die Anzahl der gedrosselten Anforderungen und die Gesamtzahl der pro Anforderungen erforderlichen Versuche.

Um die AIMD Wiederholungsstrategie zu aktivieren, müssen Sie die `fs.s3.aimd.enabled` Eigenschaft `true` in Ihrer `emrfs-site` Konfiguration wie im folgenden Beispiel auf einstellen.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.aimd.enabled": "true"
    }
  }
]
```

Weitere Informationen zum Arbeiten mit Konfigurationsobjekten finden Sie unter [Anwendungen konfigurieren](#).

Erweiterte Einstellungen für AIMD Wiederholungsversuche

Sie können die in der folgenden Tabelle aufgeführten Eigenschaften konfigurieren, um das Wiederholungsverhalten zu verfeinern, wenn Sie die AIMD Wiederholungsstrategie verwenden. Für die meisten Anwendungsfälle empfehlen wir die Verwendung der Standardeinstellung.

Erweiterte Eigenschaften der AIMD Wiederholungsstrategie

| Eigenschaft | Standardwert | Beschreibung |
|---|--------------|---|
| <code>fs.s3.aimd.increaseIncrement</code> | 0.1 | Steuert, wie schnell die Anforderungsrate steigt, |

| Eigenschaft | Standardwert | Beschreibung |
|---|--------------|---|
| | | wenn aufeinanderfolgende Anforderungen erfolgreich sind. |
| <code>fs.s3.aimd.reductionFactor</code> | 2 | Steuert, wie schnell die Anforderungsrate sinkt, wenn Amazon S3 eine 503-Antwort zurückgibt. Der Standardfaktor von 2 halbiert die Anforderungsrate. |
| <code>fs.s3.aimd.minRate</code> | 0.1 | Legt die Untergrenze für die Anforderungsrate fest, wenn Anfragen dauerhaft durch S3 gedrosselt werden. |
| <code>fs.s3.aimd.initialRate</code> | 5500 | Legt die anfängliche Anforderungsrate fest, die sich dann entsprechend den Werten ändert, die Sie für <code>fs.s3.aimd.increaseIncrement</code> und <code>fs.s3.aimd.reductionFactor</code> angeben.

Die ursprüngliche Rate wird auch für Anfragen verwendet und für GET Anfragen proportional skaliert (3500/5500). PUT |
| <code>fs.s3.aimd.adjustWindow</code> | 2 | Steuert, wie oft die Anforderungsrate angepasst wird, gemessen an der Anzahl der Antworten. |

| Eigenschaft | Standardwert | Beschreibung |
|-------------------------------------|--------------|--|
| <code>fs.s3.aimd.maxAttempts</code> | 100 | Legt die maximale Anzahl der Versuche fest, eine Anforderungen zu versuchen. |

Einen Spark-Schritt hinzufügen

Sie können Amazon EMR Steps verwenden, um Arbeiten an das auf einem EMR Cluster installierte Spark-Framework einzureichen. Weitere Informationen finden Sie unter [Schritte](#) im Amazon EMR Management Guide. In der Konsole und CLI verwenden Sie dazu einen Spark-Anwendungsschritt, der das `spark-submit` Skript als Schritt in Ihrem Namen ausführt. Mit dem API rufen Sie in einem Schritt die `spark-submit` Verwendung von `command-runner.jar` auf.

Weitere Informationen zum Senden von Anwendungen an Spark finden Sie im Thema [Bewerbungen einreichen](#) in der Apache-Spark-Dokumentation.

So senden Sie einen Spark-Schritt mithilfe der Konsole


1. Öffnen Sie die EMR Amazon-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie unter Cluster List den Namen Ihres Clusters aus.
3. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
4. Gehen Sie im Dialogfeld Add Step folgendermaßen vor:
 - Wählen Sie für Step type die Option Spark application aus.
 - Übernehmen Sie unter Name den Standardnamen ("Spark application") oder geben Sie einen neuen Namen ein.
 - Wählen Sie unter Deploy mode (Bereitstellungsmodus) den Modus Client oder Cluster aus. Der Client-Modus startet das Treiberprogramm in der Primär-Instance des Clusters, während der Cluster-Modus das Treiberprogramm im Cluster startet. Im Client-Modus erscheint die Protokollausgabe des Treibers in den Step-Logs, während im Cluster-Modus die Protokollausgabe des Treibers in den Logs für den ersten YARN Container erscheint. Weitere Informationen finden Sie unter [Übersicht über den Clustermodus](#) in der Apache-Spark-Dokumentation.

- Geben Sie die gewünschten spark-submit-Optionen an. Weitere Informationen zu den spark-submit-Optionen finden Sie unter [Starten von Anwendungen mit Spark-Submit](#).
 - Geben Sie für den Speicherort der Anwendung den lokalen URI Pfad oder den S3-Pfad der Anwendung an.
 - Lassen Sie unter Arguments (Parameter) das Feld leer.
 - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
5. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
 6. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie über der Spalte Actions (Aktionen) das Symbol Refresh (Ansicht aktualisieren) aus, um den Status zu aktualisieren.
 7. Die Ergebnisse dieses Schritts befinden sich auf der Seite Cluster-Details der EMR Amazon-Konsole neben Ihrem Schritt unter Protokolldateien, falls Sie die Protokollierung konfiguriert haben. Optional finden Sie Informationen zum Schritt im Protokoll-Bucket, den Sie beim Start des Clusters konfiguriert haben.

Um Arbeiten bei Spark einzureichen, verwenden Sie den AWS CLI

Senden Sie einen Schritt, wenn Sie den Cluster erstellen, oder verwenden Sie den Unterbefehl `aws emr add-steps` in einem vorhandenen Cluster.

1. Verwenden Sie `create-cluster`, wie im folgenden Beispiel gezeigt.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Alternativ können Sie auch `command-runner.jar` verwenden, wie im folgenden Beispiel gezeigt.

```
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

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

2. Alternativ können Sie einem bereits ausgeführten Cluster Schritte hinzufügen. Verwenden Sie `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]
```

Alternativ können Sie auch `command-runner.jar` verwenden, wie im folgenden Beispiel gezeigt.

```
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]
```

Um Arbeiten bei Spark einzureichen, verwenden Sie SDK für Java

1. Das folgende Beispiel zeigt, wie Sie einen Schritt zu einem Cluster mit Spark mittels Java hinzufügt.

```
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. Untersuchen Sie die Protokolle für den Schritt, um die Ergebnisse einzusehen. Sie können dies tun, indem Sie die Protokollierung in der AWS Management Console aktivieren, indem Sie die Schritte wählen, Ihren Schritt auswählen und dann für Protokolldateien entweder `stdout` oder `stderr` wählen. Wählen Sie `View Logs` aus, um eine Ansicht der verfügbaren Protokolle zu erhalten.

Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen

Vielleicht möchten Sie die standardmäßigen Spark-Konfigurationswerte je nach Anwendung überschreiben. Sie können dies tun, wenn Sie Anwendungen mithilfe eines Schritts senden, wodurch im Wesentlichen Optionen an `spark-submit` übergeben werden. Sie möchten z. B. den Speicher ändern, der einem Executor-Prozess zugeordnet wurde. Dazu modifizieren Sie `spark.executor.memory`. Dann geben Sie den Schalter `--executor-memory` mit einem Argument wie dem folgenden an:

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Auf ähnliche Weise können Sie `--executor-cores` und `--driver-memory` optimieren. In einem Schritt würden Sie die folgenden Argumente bereitstellen:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Sie können mithilfe der Option `--conf` auch Einstellungen optimieren, die eventuell über keinen integrierten Schalter verfügen. Weitere Informationen zu anderen optimierbaren Einstellungen finden Sie im Thema [Dynamisches Laden von Spark-Eigenschaften](#) in der Apache-Spark-Dokumentation.

Anzeigen des Spark-Anwendungsverlaufs

Sie können die Details der Benutzeroberfläche von Spark, YARN Anwendung und Tez mithilfe der Registerkarte Anwendungsbenutzeroberflächen auf der Detailseite eines Clusters in der Konsole anzeigen. Die EMR Benutzeroberflächen (UI) von Amazon erleichtern Ihnen die Fehlerbehebung und Analyse aktiver Jobs und des Jobverlaufs.

Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Amazon EMR Management Guide.

Greifen Sie auf das Spark-Web zu UIs

Sie können das Spark-Web UIs aufrufen, indem Sie den Verfahren zum Erstellen eines SSH Tunnels oder zum Erstellen eines Proxys im Abschnitt [Connect to the Cluster](#) im Amazon EMR Management Guide folgen und dann YARN ResourceManager zu Ihrem Cluster navigieren. Wählen Sie den Link unter Tracking UI für Ihre Anwendung aus. Wenn Ihre Anwendung läuft, sehen ApplicationMaster Sie. Auf diese Weise gelangen Sie zur Web-Benutzerschnittstelle der Spark-Anwendung auf Port 20888, wo auch immer sich der Treiber befindet. Der Treiber befindet sich möglicherweise auf dem primären Knoten des Clusters, wenn Sie ihn im YARN Client-Modus ausführen. Wenn Sie eine Anwendung im YARN Clustermodus ausführen, befindet sich der Treiber im Ordner ApplicationMaster für die Anwendung auf dem Cluster. Wenn Ihre Anwendung abgeschlossen ist, wird Verlauf angezeigt, der Sie zur HistoryServer Spark-UI-Portnummer 18080 des Primärknotens des EMR Clusters weiterleitet. Dies gilt für Anwendungen, die bereits abgeschlossen sind. Sie können auch direkt zur HistoryServer Spark-Benutzeroberfläche unter `http://navigierenmaster-public-dns-name:18080/`.

Mit EMR Amazon-Version 5.25.0 und höher können Sie von der Konsole aus auf die Benutzeroberfläche des Spark-History-Servers zugreifen, ohne einen Web-Proxy über eine SSH

Verbindung einrichten zu müssen. Weitere Informationen finden Sie unter [Anzeigen von persistenten Anwendungsbenuzoberflächen](#).

Verwenden des Amazon Kinesis Data Streams-Connectors für strukturiertes Streaming mit Spark

Die EMR Amazon-Versionen 7.1.0 und höher enthalten einen strukturierten Spark-Streaming-Konnektor für Amazon Kinesis Data Streams im Release-Image. Mit diesem Connector können Sie Spark on Amazon verwenden, EMR um Daten zu verarbeiten, die in Amazon Kinesis Data Streams gespeichert sind. Der Connector unterstützt sowohl Verbrauchertypen GetRecords (gemeinsamer Durchsatz) als auch SubscribeToShard (erweiterter Fan-Out). Diese Integration basiert auf dem [spark-sql-kinesis-connector](#). Einzelheiten zu den ersten Schritten mit der Verwendung des Connectors finden Sie unter [README](#).

Das folgende Beispiel zeigt, wie Sie den Connector verwenden, um eine Spark-Anwendung mit Amazon zu starten. EMR

```
spark-submit my_kinesis_streaming_script.py
```

Verwenden der Amazon Redshift Redshift-Integration für Apache Spark mit Amazon EMR

Mit EMR Amazon-Version 6.4.0 und höher enthält jedes Release-Image einen Connector zwischen [Apache Spark](#) und Amazon Redshift. Mit diesem Connector können Sie Spark on Amazon verwenden EMR, um in Amazon Redshift gespeicherte Daten zu verarbeiten. Für die EMR Amazon-Versionen 6.4.0 bis 6.8.0 basiert die Integration auf dem [spark-redshiftOpen-Source-Connector](#). Für EMR Amazon-Versionen 6.9.0 und höher wurde die [Amazon Redshift-Integration für Apache Spark](#) von der Community-Version auf eine native Integration migriert.

Themen

- [Starten einer Spark-Anwendung mithilfe der Amazon-Redshift-Integration für Apache Spark](#)
- [Mit der Amazon-Redshift-Integration für Apache Spark authentifizieren](#)
- [Lesen und Schreiben von und zu Amazon Redshift](#)
- [Überlegungen und Einschränkungen bei der Verwendung des Spark-Connectors](#)

Starten einer Spark-Anwendung mithilfe der Amazon-Redshift-Integration für Apache Spark

Für die EMR Amazon-Versionen 6.4 bis 6.9 müssen Sie die `--packages` Option `--jars` oder verwenden, um anzugeben, welche der folgenden JAR Dateien Sie verwenden möchten. Die `--jars` Option gibt Abhängigkeiten an, die lokalHDFS, in oder mithilfe von HTTP /S gespeichert sind. Weitere von der `--jars` Option unterstützte Dateispeicherorte finden Sie unter [Erweitertes Abhängigkeitsmanagement](#) in der Spark-Dokumentation. Die `--packages`-Option spezifiziert Abhängigkeiten, die im öffentlichen Maven-Repository gespeichert sind.

- `spark-redshift.jar`
- `spark-avro.jar`
- `RedshiftJDBC.jar`
- `minimal-json.jar`

EMRAmazon-Versionen 6.10.0 und höher erfordern die `minimal-json.jar` Abhängigkeit nicht und installieren die anderen Abhängigkeiten standardmäßig automatisch in jedem Cluster. Die folgenden Beispiele zeigen, wie Sie eine Spark-Anwendung mit der Amazon-Redshift-Integration für Apache Spark starten.

Amazon EMR 6.10.0 +

Das folgende Beispiel zeigt, wie eine Spark-Anwendung mit dem `spark-redshift` Connector mit EMR Amazon-Versionen 6.10 und höher gestartet wird.

```
spark-submit my_script.py
```

Amazon EMR 6.4.0 - 6.9.x

Um eine Spark-Anwendung mit dem `spark-redshift` Connector auf den EMR Amazon-Versionen 6.4 bis 6.9 zu starten, müssen Sie die `--packages` Option `--jars` oder verwenden, wie das folgende Beispiel zeigt. Beachten Sie, dass die mit der `--jars` Option aufgeführten Pfade die Standardpfade für die JAR Dateien sind.

```
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
```

Mit der Amazon-Redshift-Integration für Apache Spark authentifizieren

Wird verwendet AWS Secrets Manager , um Anmeldeinformationen abzurufen und eine Verbindung zu Amazon Redshift herzustellen

Das folgende Codebeispiel zeigt, wie Sie Anmeldeinformationen abrufen können AWS Secrets Manager , um eine Verbindung zu einem Amazon Redshift Redshift-Cluster mit der PySpark Schnittstelle für Apache Spark in Python herzustellen.

```
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()
```

Wird verwendet IAM, um Anmeldeinformationen abzurufen und eine Verbindung zu Amazon Redshift herzustellen

Sie können den von Amazon Redshift bereitgestellten JDBC Version 2-Treiber verwenden, um über den Spark-Connector eine Verbindung zu Amazon Redshift herzustellen. Um AWS Identity and Access Management (IAM) zu verwenden, [konfigurieren Sie Ihr JDBC URL System so, dass es die Authentifizierung](#) verwendet. IAM Um eine Verbindung zu einem Redshift-Cluster von Amazon herzustellen EMR, müssen Sie Ihrer IAM Rolle die Erlaubnis erteilen, temporäre IAM Anmeldeinformationen abzurufen. Weisen Sie Ihrer IAM Rolle die folgenden Berechtigungen zu, damit sie Anmeldeinformationen abrufen und Amazon S3 S3-Operationen ausführen kann.

- [Redshift: GetClusterCredentials](#) (für bereitgestellte Amazon Redshift Redshift-Cluster)
- [Redshift: DescribeClusters](#) (für bereitgestellte Amazon Redshift Redshift-Cluster)
- [Redshift: GetWorkgroup](#) (für serverlose Amazon Redshift Redshift-Arbeitsgruppen)
- [Redshift: GetCredentials](#) (für serverlose Amazon Redshift Redshift-Arbeitsgruppen)
- [s3: GetBucket](#)
- [s3: GetBucketLocation](#)
- [s3: GetObject](#)
- [s3: PutObject](#)
- [s3: GetBucketLifecycleConfiguration](#)

Weitere Informationen zu `GetClusterCredentials` finden Sie unter [Ressourcenrichtlinien für GetClusterCredentials](#).

Sie müssen auch sicherstellen, dass Amazon Redshift die IAM Rolle während des COPY UNLOAD Betriebs übernehmen kann.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "redshift.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```



```
]
}
```

Das folgende Beispiel verwendet die IAM Authentifizierung zwischen Spark und 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()
```

Lesen und Schreiben von und zu Amazon Redshift

Die folgenden Codebeispiele dienen PySpark zum Lesen und Schreiben von Beispieldaten aus und in eine Amazon Redshift Redshift-Datenbank mit Datenquelle API und unter Verwendung von SparkSQL.

Data source API

Wird verwendet PySpark , um Beispieldaten aus und in eine Amazon Redshift Redshift-Datenbank mit Datenquelle API zu lesen und zu schreiben.

```
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

Wird verwendet PySpark , um Beispieldaten mit Spark SQL aus und in eine Amazon Redshift Redshift-Datenbank zu lesen und zu schreiben.

```
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()
```

Überlegungen und Einschränkungen bei der Verwendung des Spark-Connectors

- Wir empfehlen, dass Sie die JDBC Verbindung von Spark auf Amazon EMR zu Amazon Redshift aktivieren. SSL
- Als bewährte Methode empfehlen wir Ihnen, die Anmeldeinformationen für den Amazon-Redshift-Cluster in AWS Secrets Manager zu verwalten. Ein Beispiel finden Sie [unter Verwenden AWS Secrets Manager zum Abrufen von Anmeldeinformationen für die Verbindung mit Amazon Redshift](#).
- Wir empfehlen, dass Sie eine IAM Rolle mit dem Parameter `aws_iam_role` für den Amazon Redshift Redshift-Authentifizierungsparameter übergeben.
- Das `tempdir` URI zeigt auf einen Amazon S3 S3-Standort. Dieses temporäre Verzeichnis wird nicht automatisch bereinigt und kann zusätzliche Kosten verursachen.
- Beachten Sie die folgenden Empfehlungen für Amazon Redshift:
 - Wir empfehlen, den öffentlichen Zugriff auf den Amazon-Redshift-Cluster zu blockieren.
 - Wir empfehlen, die [Amazon-Redshift-Auditprotokollierung](#) zu aktivieren.
 - Wir empfehlen Ihnen die [Amazon-Redshift-Verschlüsselung im Ruhezustand](#) zu aktivieren.
- Beachten Sie die folgenden Empfehlungen für Amazon S3:
 - Wir empfehlen Ihnen [den öffentlichen Zugriff auf Amazon-S3-Buckets zu blockieren](#).
 - Wir empfehlen die Verwendung der [serverseitigen Amazon-S3-Verschlüsselung](#), um die verwendeten Amazon-S3-Buckets zu verschlüsseln.

- Wir empfehlen, die [Lebenszyklusrichtlinien für Amazon S3](#) zu verwenden, um die Aufbewahrungsregeln für den Amazon-S3-Bucket zu definieren.
- Amazon überprüft EMR immer Code, der aus Open Source in das Bild importiert wurde. Aus Sicherheitsgründen unterstützen wir die folgenden Authentifizierungsmethoden von Spark für Amazon S3 nicht:
 - AWS Zugriffstasten in der `hadoop-env` Konfigurationsklassifizierung festlegen
 - Kodierung von AWS Zugriffsschlüsseln in der `tempdir` URI

Weitere Informationen zum Verwenden des Konnektors und seiner unterstützten Parameter finden Sie in den folgenden Ressourcen:

- [Amazon-Redshift-Integration für Apache Spark](#) im Amazon-Redshift-Verwaltungshandbuch
- Das [spark-redshift-Community-Repository](#) auf Github

Spark-Versionsverlauf

In der folgenden Tabelle sind die Version von Spark aufgeführt, die in jeder Release-Version von Amazon enthalten ist EMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Important

[Apache Spark Version 2.3.1, verfügbar ab EMR Amazon-Version 5.16.0, adressiert CVE-2018-8024 und -2018-1334. CVE](#) Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

Die Spark Versionsinformationen

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-7.2.0 | 3.5.1 | delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| | | hadoop-client, hadoop-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-7.1.0 | 3.5.0 | delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-7.0.0 | 3.5.0 | delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-6.15.0 | 3.4.1 | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.14.0 | 3.4.1 | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-6.13.0 | 3.4.1 | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.12.0 | 3.4.0 | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-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-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-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-6.10.1 | 3.3.1 | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.10.0 | 3.3.1 | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-6.7.0 | 3.2.1 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.36.1 | 2.4.8 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-5.35.0 | 2.4.8 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.5.0 | 3.1.2 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-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-resource-manager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-6.0.0 | 2.4.4 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.34.0 | 2.4.8 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.32.1 | 2.4.7 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.32.0 | 2.4.7 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.31.0 | 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-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-5.20.1 | 2.4.0 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.20.0 | 2.4.0 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.17.0 | 2.3.1 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.16.1 | 2.3.1 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-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-5.15.1 | 2.3.0 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-ftpfs-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-ftpfs-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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-httfs-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-httfs-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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-httfs-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-httfs-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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-httfs-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-httfs-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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-httfs-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-httfs-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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.5.2 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.5.1 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.5.0 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.4.1 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.4.0 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.3.2 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.3.1 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.3.0 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.3 | 2.0.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.2.2 | 2.0.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.1 | 2.0.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.0 | 2.0.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-5.1.1 | 2.0.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.1.0 | 2.0.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.0.3 | 2.0.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-5.0.2 | 2.0.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-4.9.6 | 1.6.3 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.9.5 | 1.6.3 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.9.4 | 1.6.3 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.4 | 1.6.3 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.3 | 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-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-4.8.2 | 1.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.1 | 1.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.0 | 1.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-4.7.4 | 1.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.7.3 | 1.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.7.2 | 1.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-4.6.0 | 1.6.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht
tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource
manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.5.0 | 1.6.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht
tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource
manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.4.0 | 1.6.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht
tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource
manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|---|
| emr-4.3.0 | 1.6.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht
tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource
manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.2.0 | 1.5.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht
tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource
manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.1.0 | 1.5.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht
tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource
manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon EMR Release-Etikett | Spark-Version | Mit Spark installierte Komponenten |
|----------------------------|---------------|--|
| emr-4.0.0 | 1.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

Apache Sqoop

Apache Sqoop ist ein Tool für die Übertragung von Daten zwischen Amazon S3, Hadoop und HDFS RDBMS Datenbanken. Weitere Informationen finden Sie auf der [Apache Sqoop-Website](#). Sqoop ist in den EMR Amazon-Versionen 5.0.0 und höher enthalten. Frühere Versionen enthalten Sqoop als Sandbox-Anwendung. Weitere Informationen finden Sie unter [Amazon EMR 4.x-Release-Versionen](#).

Themen

- [Die Sqoop-Versionsinformationen](#)
- [Überlegungen zu Sqoop auf Amazon EMR](#)
- [Versionsverlauf von Sqoop](#)

Die Sqoop-Versionsinformationen

Sqoop Version für 7.2.0

In der folgenden Tabelle sind die Version von Sqoop aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Sqoop EMR installiert.

[Die Version der Komponenten, die in dieser Version mit Sqoop installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Sqoop Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|---------------------------|---------------|--|
| 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- |

| EMRAmazon-Freigabeetikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|---------------------------|---------------|---|
| | | timeline-server, mariadb-server, sqoop-client |

Sqoop-Version für 6.15.0

In der folgenden Tabelle sind die Version von Sqoop aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Sqoop EMR installiert.

Die Version der Komponenten, die mit Sqoop in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Sqoop-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|---------------------------|---------------|---|
| 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 |

Sqoop Version für 5.36.2

In der folgenden Tabelle sind die Version von Sqoop aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon mit Sqoop EMR installiert.

[Die Version der Komponenten, die in dieser Version mit Sqoop installiert wurden, finden Sie unter Komponentenversionen der Version 5.36.2.](#)

Sqoop Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|---------------------------|---------------|---|
| 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 |

Überlegungen zu Sqoop auf Amazon EMR

Beachten Sie die folgenden Punkte, wenn Sie Sqoop bei Amazon EMR ausführen.

Verwenden von Sqoop mit Integration HCatalog

Sqoop auf Amazon EMR unterstützt die [Sqoop- HCatalog Integration](#). Wenn Sie Sqoop verwenden, um Ausgaben in eine HCatalog Tabelle in Amazon S3 zu schreiben, deaktivieren Sie Amazon EMR Direct Write, indem Sie die `mapred.output.direct.EmrFileSystem` Eigenschaften `mapred.output.direct.NativeS3FileSystem` und auf `false` setzen. Weitere Informationen finden Sie unter [Verwenden HCatalog](#). Sie können die Hadoop-Befehle `-D mapred.output.direct.NativeS3FileSystem=false` und `-D mapred.output.direct.EmrFileSystem=false` verwenden. Wenn Sie die Direct-Write-Funktion nicht deaktivieren, tritt kein Fehler auf, aber die Tabelle wird in Amazon S3 erstellt und es werden keine Daten geschrieben.

Unterstützung für Sqoop JDBC und Datenbanken

Standardmäßig hat Sqoop einen MariaDB- und SQL Postgre-Treiber installiert. Der für Sqoop installierte SQL Postgre-Treiber funktioniert nur für Postgre 8.4. SQL Um einen alternativen Satz von JDBC Konnektoren für Sqoop zu installieren, stellen Sie eine Verbindung zum Cluster-Masterknoten her und installieren Sie sie dort. `/usr/lib/sqoop/lib` Im Folgenden finden Sie Links für verschiedene JDBC Konnektoren:

- MariaDB: [Informationen zu MariaDB Connector/J](#).
- Postgre-TreiberSQL: [Postgre-Treiber. SQL JDBC](#)
- SQLServer: [Laden Sie den JDBC Microsoft-Treiber für SQL Server](#) herunter.
- MeinSQL: [Laden Sie Connector/J herunter](#)
- Oracle: [Holen Sie sich JDBC Oracle-Treiber und UCP aus dem Oracle Maven-Repository](#)

Die unterstützten Datenbanken für Sqoop sind unter der folgenden URL aufgeführt: [http://sqoop.apache.org/docs/*version*/SqoopUserGuide.html#supported_databases](http://sqoop.apache.org/docs/version/SqoopUserGuide.html#supported_databases), wo *version* ist die Version von Sqoop, die Sie verwenden, zum Beispiel 1.4.6. Wenn die JDBC Verbindungszeichenfolge nicht mit denen in dieser Liste übereinstimmt, müssen Sie einen Treiber angeben.

Sie können beispielsweise mit dem folgenden Befehl (für JDBC 4.1) in eine Amazon Redshift Redshift-Datenbanktabelle exportieren:

```
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
```

Sie können sowohl die MariaDB- als auch die Meine SQL Verbindungszeichenfolge verwenden, aber wenn Sie die MariaDB-Verbindungszeichenfolge angeben, müssen Sie den Treiber angeben:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username master --password Mymasterpass1
```

Wenn Sie die Secure Socket Layer-Verschlüsselung für den Zugriff auf Ihre Datenbank verwenden, müssen Sie ein JDBC URI Beispiel wie im folgenden Sqoop-Exportbeispiel verwenden:

```
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
```

Weitere Informationen zur SSL Verschlüsselung in finden Sie unter [Verwenden RDS, SSL um eine Verbindung zu einer DB-Instance zu verschlüsseln](#) im RDS Amazon-Benutzerhandbuch.

Weitere Informationen finden Sie in der [Apache Sqoop](#)-Dokumentation.

Schützen Ihres Passworts

Es gibt mehrere Methoden, wie Sie Ihr Passwort sicher weitergeben können:

Java KeyStore

Die bevorzugte Methode verschlüsselt das Passwort mit einem Java KeyStore (JKS), sodass das Passwort nicht in einem lesbaren Format gespeichert werden muss.

1. Erstellen Sie einen Passwort-Alias. Geben Sie bei entsprechender Aufforderung das Passwort ein, mit dem Sie auf die Datenbank zugreifen.

```
hadoop credential create mydb.password.alias -provider jceks://hdfs/user/root/  
mysql.password.jceks
```

2. Verwenden Sie den Passwort-Alias, um den Sqoop-Auftrag zu starten:

```
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

Sie können wie im folgenden Beispiel dargestellt den `--password-file`-Befehl verwenden, um das Passwort durch eine Datei zu übergeben:

1. Erstellen Sie eine neue Datei, die das Passwort enthält:

```
echo -n 'Mymasterpass1' > /home/hadoop/mysql-pass.password
```


2. Verwenden Sie die Datei, um den Sqoop-Auftrag zu starten:

```
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

Sie können wie im folgenden Beispiel dargestellt den -P-Befehl verwenden, um das Passwort durch eine Aufforderung zu übergeben:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --
export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username
master -P
```

Versionsverlauf von Sqoop

In der folgenden Tabelle sind die Version von Sqoop aufgeführt, die in jeder Release-Version von Amazon enthalten ist, zusammen mit den mit der Anwendung installierten Komponenten. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Sqoop-Versionsinformationen

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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- |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| | | 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|---|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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-https-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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-https-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.2.1 | 1.4.6 | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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 |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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-https-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon EMR Release-Etikett | Sqoop-Version | Mit Sqoop installierte Komponenten |
|----------------------------|---------------|--|
| 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-resource-manager, mysql-server, sqoop-client |

TensorFlow

TensorFlow ist eine symbolische Open-Source-Mathematikbibliothek für maschinelle Intelligenz und Deep-Learning-Anwendungen. Weitere Informationen finden Sie [TensorFlow auf der Website](#). TensorFlow ist mit der EMR Amazon-Release-Version 5.17.0 und höher verfügbar.

In der folgenden Tabelle sind die Version von TensorFlow aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. TensorFlow

Informationen zur Version der Komponenten, mit denen TensorFlow in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

TensorFlow Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | TensorFlow Ausführung | Komponenten, die mit installiert wurden TensorFlow |
|---------------------------|-----------------------|--|
| emr-7.2.0 | 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 |

In der folgenden Tabelle sind die Version von TensorFlow aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. TensorFlow

Informationen zur Version der Komponenten, mit denen TensorFlow in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 6.15.0](#).

TensorFlow Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | TensorFlow Ausführung | Komponenten, die mit installiert wurden 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |

In der folgenden Tabelle sind die Version von TensorFlow aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert TensorFlow

Informationen zur Version der Komponenten, mit denen TensorFlow in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

TensorFlow Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | TensorFlow Ausführung | Komponenten, die mit installiert wurden 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |

TensorFlow Builds nach EC2 Amazon-Instanztyp

Amazon EMR verwendet je nach den Instance-Typen, die Sie für Ihren Cluster auswählen, unterschiedliche Builds der TensorFlow Bibliothek. In der folgenden Tabelle sind die Builds nach Instance-Typ gruppiert aufgelistet.

| EC2Instance-Typen | TensorFlow bauen |
|-------------------|--|
| M5 und C5 | Tensorflow 1.9.0 mit Intel-Optimierung MKL |
| P2 | Tensorflow 1.9.0 mit 9.2, cu 7.1 CUDA DNN |
| P3 | Tensorflow 1.9.0 mit CUDA 9.2, Cu 7.1, 2.2.13 DNN NCCL

Nvidia NCCL ist nur auf P3-Instances verfügbar
. Endbenutzer-Lizenzvereinbarung (EULA):
Durch die Verwendung von Nvidia-Komponenten bei Amazon EMR erklären Sie sich mit den im Produkt aufgeführten Bedingungen einverstanden EULA. |
| Alle anderen | TensorFlow 1.9.0 |

Sicherheit

Zusätzlich zu den Anweisungen [unter TensorFlow Sichere Nutzung](#) empfehlen wir Ihnen, Ihren Cluster in einem privaten Subnetz zu starten, um den Zugriff auf vertrauenswürdige Quellen zu beschränken. Weitere Informationen finden Sie unter [VPC Amazon-Optionen](#) im Amazon EMR Management Guide.

Verwenden TensorBoard

TensorBoard ist eine Suite von Visualisierungstools für TensorFlow Programme. Weitere Informationen finden Sie unter [TensorBoard: Visualisiertes Lernen](#) auf der Tensorflow-Website.

Für die Verwendung TensorBoard mit Amazon EMR müssen Sie TensorBoard auf dem Cluster-Masterknoten beginnen.

Um Tensorboard mit Tensorflow bei Amazon zu verwenden EMR

1. Stellen Sie mithilfe von Connect zum Master-Knoten des Clusters herSSH. Weitere Informationen finden Sie unter [Connect dem Master-Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie den folgenden Befehl ein, um TensorBoard auf dem Master-Knoten zu starten. Ersetzen Sie */my/log/directory* mit dem Verzeichnis auf dem Master-Knoten, in dem Sie unter Verwendung einer Summary Writer-Operation die Übersichtsdaten gespeichert haben.

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
```

Standardmäßig verwendet der Master-Knoten den TensorBoard Port 6006 und den öffentlichen DNS Masternamen. Nach dem Start TensorBoard wird in der Befehlszeilenausgabe URL das angezeigt, mit dem eine Verbindung hergestellt werden kann TensorBoard, wie im folgenden Beispiel gezeigt:

```
TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
```

3. Richten Sie den Zugriffs auf Webschnittstellen auf dem Master-Knoten von vertrauenswürdigen Clients aus ein. Weitere Informationen finden Sie unter [Auf EMR Amazon-Clustern gehostete Weboberflächen](#) anzeigen im Amazon EMR Management Guide.
4. Öffnet TensorBoard unter `http://master-public-dns-name:6006`.

TensorFlow Verlauf der Veröffentlichung

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon TensorFlow enthalten sindEMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

TensorFlow Versionsinformationen

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-7.0.0 | 2.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| | | hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.15.0 | 2.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.14.0 | 2.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit 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 |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| emr-6.11.0 | 2.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.10.1 | 2.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.10.0 | 2.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-6.9.1 | 2.10.0 | 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-6.9.0 | 2.10.0 | 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-6.8.1 | 2.9.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-6.8.0 | 2.9.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.7.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.36.1 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.36.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.6.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.35.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| emr-6.5.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.4.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.3.1 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-6.3.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.2.1 | 2.3.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.2.0 | 2.3.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-6.1.1 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-6.0.0 | 1.14.0 | 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-5.34.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.33.1 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| emr-5.33.0 | 2.4.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.32.1 | 2.3.1 | 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, tensorflow |
| emr-5.32.0 | 2.3.1 | 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, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.31.1 | 2.1.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| emr-5.30.1 | 1.14.0 | 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-5.30.0 | 1.14.0 | 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-5.29.0 | 1.14.0 | 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 |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.28.1 | 1.14.0 | 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-5.28.0 | 1.14.0 | 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-5.27.1 | 1.14.0 | 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 |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.27.0 | 1.14.0 | 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-5.26.0 | 1.13.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.25.0 | 1.13.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.24.1 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| emr-5.23.0 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.22.0 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.21.2 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.21.1 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.21.0 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.20.1 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|--|
| emr-5.20.0 | 1.12.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.19.1 | 1.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.19.0 | 1.11.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit TensorFlow |
|----------------------------|-----------------------|---|
| emr-5.18.1 | 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-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon EMR Release-Etikett | TensorFlow Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |

Apache Tez

Apache Tez ist ein Framework, das einen komplexen gerichteten azyklischen Graphen (DAG) von Aufgaben zur Datenverarbeitung erstellt. In einigen Anwendungsfällen können Sie es als Alternative zu Hadoop MapReduce verwenden. Sie können beispielsweise Pig- und Hive-Workflows mit Hadoop ausführen MapReduce oder Tez als Ausführungs-Engine verwenden. Weitere Informationen finden Sie unter <https://tez.apache.org/>. EMRAmazon-Versionen 4.7.0 und höher enthalten Tez.

In der folgenden Tabelle sind die Version von Tez aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Tez EMR installiert.

[Die Version der Komponenten, die in dieser Version mit Tez installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Tez-Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Tez-Version | Mit Tez installierte Komponenten |
|---------------------------|-------------|---|
| 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 |

In der folgenden Tabelle sind die Version von Tez aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Tez EMR installiert.

Die Version der Komponenten, die mit Tez in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0.](#)

Tez-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Tez-Version | Mit Tez installierte Komponenten |
|---------------------------|-------------|---|
| 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 |

In der folgenden Tabelle sind die Version von Tez aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Tez EMR installiert.

[Die Version der Komponenten, die in dieser Version mit Tez installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Tez-Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | Tez-Version | Mit Tez installierte Komponenten |
|---------------------------|-------------|--|
| 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 |

Themen

- [Erstellen eines Clusters mit Tez](#)
- [Konfigurieren von Tez](#)
- [Tez-Webbenutzeroberfläche](#)
- [Timeline-Server](#)
- [Tez-Versionsverlauf](#)

Erstellen eines Clusters mit Tez

Um Tez zu installieren, wählen Sie Apache Tez als Anwendung, wenn Sie Ihren Cluster erstellen.

So erstellen Sie einen Cluster mit Tez über die Konsole

1. Navigieren Sie zur neuen EMR Amazon-Konsole und wählen Sie in der Seitennavigation die Option [Zur alten Konsole wechseln](#) aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie **Create Cluster (Cluster erstellen)** und **Go to advanced options (Zu erweiterten Optionen)** aus.
3. Wählen Sie unter **Softwarekonfiguration** das Release `emr-4.7-0` oder höher aus.
4. Wählen Sie **Tez** zusammen mit anderen Anwendungen aus, die Amazon installieren EMR soll.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend **Create cluster (Cluster erstellen)** aus.

Um einen Cluster mit Tez zu erstellen, verwenden Sie AWS CLI

- Verwenden Sie den `create-cluster`-Befehl zusammen mit der `-- applications`-Option, um Tez festzulegen. Im folgenden Beispiel wird ein Cluster mit installiertem Tez erstellt.

Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Konfigurieren von Tez

Sie können Tez anpassen, indem Sie mit der Konfigurationsklassifizierung `tez-site` Einstellungen in der Konfigurationsdatei `tez-site.xml` konfigurieren. Weitere Informationen finden Sie [TezConfiguration](#) in der Apache Tez-Dokumentation. Verwenden Sie entsprechend die `hive-site-` und `pig-properties`-Konfigurationsklassifizierungen, um zu ändern, dass Hive oder Pig die Tez-Ausführungs-Engine verwendet. Beispiele werden unten angezeigt.

Beispielkonfiguration

Example Beispiel: Anpassen der Tez-Stammprotokollierungsebene und Festlegen von Tez als Ausführungs-Engine für Hive und Pig

Mit dem unten angezeigten `create-cluster`-Beispielbefehl wird ein Cluster mit installierten Tez, Hive und Pig erstellt. Der Befehl verweist auf eine Datei die in Amazon S3 `myConfig.json` gespeichert ist und die Eigenschaften für die `tez-site`-Klassifizierung angibt, die `tez.am.log.level` auf `DEBUG` setzt und die Ausführungs-Engine für Hive und Pig mit den Konfigurationsklassifizierungen `hive-site` und `pig-properties` auf Tez festlegt.

Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
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
```

Beispielinhalte von `myConfig.json` werden unten angezeigt.

```
[
  {
    "Classification": "tez-site",
    "Properties": {
      "tez.am.log.level": "DEBUG"
    }
  },
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.execution.engine": "tez"
    }
  },
  {
    "Classification": "pig-properties",
    "Properties": {
      "exectype": "tez"
    }
  }
]
```

Note

Mit EMR Amazon-Version 5.21.0 und höher können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifizierungen für jede Instance-Gruppe in einem laufenden Cluster angeben. Sie tun dies, indem Sie die EMR Amazon-Konsole, die AWS Command Line Interface (AWS CLI) oder die verwenden AWS SDK. Weitere Informationen finden Sie unter [Angaben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

Asynchrone Öffnung von Tez-Splits

Wenn der Tabellenpfad eine große Anzahl kleiner Dateien enthält und eine Abfrage versucht, sie alle zu lesen, wird jede kleine Datei, die jeder einzelnen Aufteilung entspricht, zu einem gruppierten Tez-Split zusammengefasst. Ein einzelner Mapper verarbeitet dann den einzelnen gruppierten Tez-Split. Da die Ausführung synchron ist, wird jeder einzelne Split im Rahmen des gruppierten Splits einzeln verarbeitet. Dafür ist erforderlich, dass `RecordReader`-Objekte die Splits synchron verarbeiten.

| Name | Klassifizierung | Beschreibung |
|---|-----------------------|---|
| <code>tez.grouping.split
.init.threads</code> | <code>tez-site</code> | Gibt die Anzahl der Daemon-Threads an, die Tez verwendet, um die <code>RecordReaders</code> vorab zu initiieren und Splits zu öffnen. Für ACID-Tabellen ist der unterstützte Höchstwert von <code>tez.grouping.split.init.threads</code> 1. |
| <code>tez.grouping.split
.init.recordreaders</code> | <code>tez-site</code> | Gibt die Anzahl an <code>RecordReaders</code> an, die von den Daemon-Threads vorinitialisiert werden sollen. Dies kann hilfreich sein, wenn ein gruppierter Tez-Split eine große Anzahl von <code>InputSplits</code> enthält. Die Initialisierung von <code>RecordReaders</code> zur Verarbeitung dieser Eingabe-Splits kann asynchron mit Daemon-Threads anstelle einer sequentiellen Verarbeitung erfolgen. |

Benchmarking für das asynchrone Öffnen von Tez-Splits

Wir haben die folgenden Umgebungen und Konfigurationen für das Benchmarking der asynchronen Split-Öffnungsfunktion von Tez verwendet:

- Benchmark-Umgebung — EMR Amazon-Cluster mit einem primären Knoten, der verwendet `m5.16xlarge`, und 16 Kernknoten, die verwenden `m5.16xlarge`.

- Benchmark-Konfigurationen – Um das Szenario für das Benchmarking zu simulieren, bei dem sich eine große Anzahl von Eingabe-Splits in einem einzigen gruppierten Tez-Split befinden, ist `tez.grouping.split-count` auf 1 eingestellt.
- Für das Benchmarking verwendete Tabelle – Die Tabelle enthält 200 Partitionen, wobei jede Partition eine einzelne Datei enthält. Der Benchmark wird durchgeführt, wenn diese Tabelle CSV-Dateien enthält und wenn diese Tabelle Parquet-Dateien enthält. Hive-Abfrage für das Benchmarking: zehnmal `SELECT COUNT(*)` aus der Tabelle, und es wird die durchschnittliche Laufzeit berechnet.
- Konfigurationen zur Aktivierung des asynchronen geteilten Öffnens von Tez – wie folgt:
 - `tez.grouping.split.init.threads = 4`
 - `tez.grouping.split.init.recordreaders = 10`

| Datensatz | Feature deaktiviert (Basisversion) | Feature aktiviert | Verbesserung |
|-------------------|------------------------------------|-------------------|--------------|
| CSV-Datensatz | 90,26 Sekunden | 79,20 Sekunden | 12,25% |
| Parquet-Datensatz | 54,67 Sekunden | 42,23 Sekunden | 22,75% |

Tez-Webbenutzeroberfläche

Tez verfügt über eine eigene Webbenutzeroberfläche. Informationen zur Webbenutzeroberfläche finden Sie im Folgenden URL.

```
http://masterDNS:8080/tez-ui
```

Um die Registerkarte Hive-Abfragen auf der Tez-Weboberfläche zu aktivieren, stellen Sie die folgende Konfiguration ein.

```
[
  {
    "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"
    }
  }
]
```

```

    }
  }
]
```

Sie können die Details zu Tez, Spark und der YARN Anwendungsbenuzoberfläche auch mithilfe von Links auf der Registerkarte Anwendungsbenuzoberflächen auf der Detailseite eines Clusters in der Konsole anzeigen. Die EMR Benutzeroberflächen (UI) von Amazon werden außerhalb des Clusters gehostet und sind verfügbar, nachdem der Cluster beendet wurde. Sie erfordern nicht, dass Sie eine SSH Verbindung oder einen Webproxy einrichten, was es Ihnen erleichtert, Fehler zu beheben und aktive Jobs und den Jobverlauf zu analysieren.

Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Amazon EMR Management Guide.

Timeline-Server

Der YARN Timeline Server ist so konfiguriert, dass er läuft, wenn Tez installiert ist. Um Jobs anzuzeigen, die über Tez oder MapReduce Execution Engines mit dem Timeline Server eingereicht wurden, rufen Sie die Weboberfläche mit dem auf. URL `http://master-public-DNS:8188`
 Weitere Informationen finden Sie unter [Auf EMR Amazon-Clustern gehostete Weboberflächen anzeigen](#) im Amazon EMR Management Guide.

Tez-Versionsverlauf

In der folgenden Tabelle sind die in jeder Release-Version von Amazon EMR enthaltenen Tez-Versionen sowie die mit der Anwendung installierten Komponenten aufgeführt. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Die Tez-Versionsinformationen

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| emr-7.2.0 | 0.10.2 | emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop- |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| | | 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 |
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|---|
| 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn |

| Amazon EMR Release-Etikett | Tez-Version | Mit Tez installierte Komponenten |
|----------------------------|-------------|--|
| 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 |

Tez-Versionshinweise nach Version

Themen

- [Versionshinweise zu Amazon EMR 7.1.0 — Tez](#)
- [Amazon EMR 6.15.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.14.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.13.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.12.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.11.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.10.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.9.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.8.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.7.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.6.0 — Tez-Versionshinweise](#)

Versionshinweise zu Amazon EMR 7.1.0 — Tez

Amazon EMR 7.1.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------------|---|
| Fehlerbehebung | TEZ-4394 — Netty4 ShuffleHandler: sollte 1 Boss-Thread verwenden (#256) |
| Upgrade | Führt ein Upgrade von Netty auf 4.1.100.Final durch. |
| Upgrade | Führt ein Upgrade von Jetty auf 9.4.53.v2 0231009 durch. |

Amazon EMR 6.15.0 — Tez-Versionshinweise

Amazon EMR 6.15.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------|--|
| Funktion | TEZ-4397 : Open Tez Input teilt sich asynchron |

| Typ | Beschreibung |
|---------|--|
| Upgrade | TEZ-4493 : Aktualisieren Sie Apache Hadoop auf 3.3.6 |

Amazon EMR 6.15.0 — Tez-Funktionen

- [Asynchrones geteiltes Öffnen von Tez](#) — Amazon EMR 6.15.0 führt Konfigurationen ein, die Sie angeben können, um die Eingabe-Splits in einem gruppierten Tez-Split asynchron zu öffnen. Die Funktion wurde von -4397 initiiert, wies jedoch [TEZRegressionen](#) in Hive auf. OSS Amazon EMR Hive hat die Regressionen und zusätzliche Fehler in der Hive-Tabelle behoben. ACID Diese Verbesserung führt zu einer schnelleren Leistung von Leseabfragen, wenn ein einzelner gruppierter Tez-Split eine große Anzahl von Eingabe-Splits enthält. Weitere Informationen finden Sie unter [Asynchrone Öffnung von Tez-Splits](#).

Amazon EMR 6.14.0 — Tez-Versionshinweise

Amazon EMR 6.14.0 — Tez ändert sich

| Typ | Beschreibung |
|--------------|--|
| Verbesserung | Aktualisieren Sie die TLS Version in Tez auf 1.2 |

Amazon EMR 6.13.0 — Tez-Versionshinweise

Amazon EMR 6.13.0 — Tez ändert sich

| Typ | Beschreibung |
|----------------|---|
| Fehlerbehebung | REVERT TEZ-4295 : Daten konnten nicht dekomprimiert werden. Die Pufferlänge ist zu klein. |
| Fehlerbehebung | REVERT TEZ-4302 : drin mit. NullPointerException CodecUtils GzipCodec |

| Typ | Beschreibung |
|----------------|--|
| Fehlerbehebung | REVERT TEZ-4234 : Der Kompressor kann <code>Buffer.limit</code> bei Überschreitung <code>IllegalArgumentException</code> der Kapazität eine Störung verursachen. |
| Fehlerbehebung | REVERT TEZ-4135 : Verbessert die Speicherzuweisung bei der Ausführung von Lesevorgängen im Speicher. |

Amazon EMR 6.12.0 — Tez-Versionshinweise

Amazon EMR 6.12.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------------|--|
| Verbesserung | Support für JDK 11 und JDK 17 Runtime hinzugefügt |
| Fehlerbehebung | TEZ-4492 : Aktualisieren Sie <code>Bowerrc</code> , um den <code>bower.herokuapp</code> -Spiegel zu verwenden, um das <code>Bower-Registry _ Problem</code> zu vermeiden (-2608) CERT EXPIRE BOWER |
| Upgrade | Surefire wurde auf 3.0.0-M7 aktualisiert |

Amazon EMR 6.11.0 — Tez-Versionshinweise

Amazon EMR 6.11.0 — Tez-Änderungen

| Typ | Beschreibung |
|--------|--|
| Fehler | Ein ungültiger Übergang zum Scheitelpunktstatus bei der Bereinigung von Shuffle-Daten auf Scheitelpunktebene wurde behoben |

| Typ | Beschreibung |
|--------------|--|
| Fehler | Fehler behoben DAG oder die Bereinigung von Shuffle-Daten auf Scheitelpunktebene funktioniert nicht |
| Verbesserung | Aktiviert standardmäßig <code>tez.am.dag.cleanup.on.completion</code> , um die Shuffle-Daten von Completed zu löschen DAGs |

Amazon EMR 6.10.0 — Tez-Versionshinweise

Amazon EMR 6.10.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------|---|
| Funktion | <code>tez.runtime.transfer.data-via-events.enabled</code> standardmäßig aktivieren |
| Backport | TEZ-4450 : Der Fehler beim Abrufen von Shuffle-Daten, wenn Shuffle-Daten im Rahmen von Datenverschiebungsereignissen übertragen wurden, wurde behoben |
| Backport | TEZ-4460 : Der Lese-Timeout-Fehler beim Abrufen von Shuffle-Daten vom Tez Shuffle Handler wurde behoben |
| Backport | TEZ-4455 : Pipeline für bessere Debuggbarkeit hinzufügen LoggingHandler ShuffleHandler |
| Fehler | Behebt, dass die Tez-Aufgabe zeitweise hängen bleibt, wenn das Preemptive der Aufgabe aktiviert ist |

Amazon EMR 6.9.0 — Tez-Versionshinweise

Amazon EMR 6.9.0 — Tez-Änderungen

| Typ | Beschreibung |
|---------|---|
| Upgrade | Tez wurde auf 0.10.2 aktualisiert. Weitere Informationen finden Sie im Änderungsprotokoll für Apache Tez 0.10.2 . |
| Upgrade | Upgrade auf Hadoop 3.3.3. |
| Fehler | Standardmäßig deaktiviert aufgrund tez.runtime.transfer.data-via-events.enabled von -4450. TEZ |

Amazon EMR 6.8.0 — Tez-Versionshinweise

Amazon EMR 6.8.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------|---|
| Backport | TEZ-3363 : Löscht Zwischendaten auf Scheitelpunktebene für Shuffle Handler |
| Backport | TEZ-4129 : Löscht Daten für Zwischenversuche bei fehlgeschlagenen Versuchen für Shuffle Handler |
| Backport | TEZ-4430 : Problem behoben, dass die tez.task.launch.cmd-opts-Eigenschaft nicht funktionierte |

Amazon EMR 6.7.0 — Tez-Versionshinweise

Amazon EMR 6.7.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------|---|
| Backport | TEZ-4403 : Version auf 1.7.36 aktualisieren SLF4J |
| Backport | TEZ-4405 : Ersetze log4j 1.x durch reload4j |
| Backport | TEZ-4411: Tez-Build-Fehler : .js nicht gefunden FileSaver |

Amazon EMR 6.6.0 — Tez-Versionshinweise

Amazon EMR 6.6.0 — Tez-Änderungen

| Typ | Beschreibung |
|----------|---|
| Backport | TEZ-3918 : Es wurde behoben, dass die Eigenschaft tez.task.log.level nicht funktionierte. |
| Backport | TEZ-4353 : Aktualisieren Sie commons-io auf 2.8.0. |
| Backport | TEZ-4114 : Entfernt die direkte Jetty-Abhängigkeit von Tez. |
| Backport | TEZ-4323 : Jetty-Jars wurden mit -4114 aus dem Paket dist entfernt. TEZ |

Apache Zeppelin

Verwenden Sie Apache Zeppelin als Notizbuch für eine interaktive Datenauswertung. Weitere Informationen zu Zeppelin finden Sie unter <https://zeppelin.apache.org/>. Zeppelin ist in den EMR Amazon-Versionen 5.0.0 und höher enthalten. Frühere Versionen enthalten Zeppelin als Sandbox-Anwendung. Weitere Informationen finden Sie unter [Amazon EMR 4.x-Release-Versionen](#).

Um auf die Zeppelin-Weboberfläche zuzugreifen, richten Sie einen SSH Tunnel zum Master-Knoten und eine Proxyverbindung ein. Weitere Informationen finden Sie unter [Auf Clustern gehostete Weboberflächen anzeigen](#). EMR

In der folgenden Tabelle sind die Version von Zeppelin aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Zeppelin EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Zeppelin installiert wurden, finden Sie unter Komponentenversionen der Version 7.2.0.](#)

Zeppelin-Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|---------------------------|------------------|---|
| emr-7.2.0 | Zeppelin 0.10.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

In der folgenden Tabelle sind die Version von Zeppelin aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Zeppelin EMR installiert.

Die Version der Komponenten, die mit Zeppelin in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Zeppelin-Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|---------------------------|------------------|---|
| 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 |

In der folgenden Tabelle sind die Version von Zeppelin aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, die Amazon zusammen mit Zeppelin EMR installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Zeppelin installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Zeppelin-Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|---------------------------|------------------|--|
| emr-5.36.2 | Zeppelin 0.10.0 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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 |

Themen

- [Überlegungen zur Verwendung von Zeppelin bei Amazon EMR](#)
- [Zeppelin-Versionshistorie](#)

Überlegungen zur Verwendung von Zeppelin bei Amazon EMR

- Stellen Sie mithilfe derselben [SSTunneling-Methode](#) eine Connect zu Zeppelin her, um eine Verbindung zu anderen Webservern auf dem Master-Knoten herzustellen. Der Zeppelin-Server befindet sich an Port 8890.
- [Zeppelin auf Amazon EMR veröffentlicht Versionen 5.0.0 und höher und unterstützt die Shiro-Authentifizierung.](#)
- Zeppelin auf Amazon EMR veröffentlicht die Versionen 5.8.0 und höher und unterstützt die Verwendung von AWS Glue Data Catalog als Metastore für Spark. SQL Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Spark SQL.](#)

- Zeppelin verwendet einige der in der `spark-defaults.conf` Konfigurationsdatei Ihres Clusters definierten Einstellungen nicht, obwohl es anweist, Executoren dynamisch YARN zuzuweisen, falls Sie dies eingestellt haben. `spark.dynamicAllocation.enabled true` Sie müssen Executor-Einstellungen, z. B. Arbeitsspeicher und Prozessorkerne, über die Zeppelin-Registerkarte Interpreter festlegen. Starten Sie dann den Interpreter neu, um die Änderungen zu übernehmen.
- EMRAmazon-Versionen 6.10.0 und höher unterstützen die Apache Zeppelin-Integration mit Apache Flink. Weitere Informationen finden Sie unter [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#).
- Zeppelin auf Amazon EMR unterstützt den SparkR-Interpreter nicht.

Zeppelin-Versionshistorie

In der folgenden Tabelle sind die Version von Zeppelin aufgeführt, die in jeder Release-Version von Amazon enthalten istEMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

Zeppelin-Versionsinformationen

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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 |
| 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-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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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-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-5.20.1 | 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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-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-5.19.1 | 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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-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-5.18.1 | 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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-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-5.12.3 | 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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-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-5.12.1 | 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.12.0 | 0.7.3 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.11.4 | 0.7.3 | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.5.2 | 0.7.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.5.1 | 0.7.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.5.0 | 0.7.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.4.1 | 0.7.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.4.0 | 0.7.0 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.3.2 | 0.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.2.3 | 0.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.2.2 | 0.6.2 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|---|
| 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 |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| 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-resource-manager, 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-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.0.3 | 0.6.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.0.2 | 0.6.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon EMR Release-Etikett | Zeppelin-Version | Mit Zeppelin installierte Komponenten |
|----------------------------|------------------|--|
| emr-5.0.1 | 0.6.1 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| 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-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

Apache ZooKeeper

Apache ZooKeeper ist ein zentraler Dienst zur Verwaltung von Konfigurationsinformationen, zur Benennung, zur Bereitstellung verteilter Synchronisation und zur Bereitstellung von Gruppendiensten. Weitere Informationen zu ZooKeeper finden Sie unter <http://zookeeper.apache.org/>.

In der folgenden Tabelle sind die Version von ZooKeeper aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. ZooKeeper

Informationen zur Version der Komponenten, mit denen ZooKeeper in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 7.2.0](#).

ZooKeeper Versionsinformationen für emr-7.2.0

| EMRAmazon-Freigabeetikett | ZooKeeper Ausführung | Komponenten, die mit installiert wurden 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 |

In der folgenden Tabelle sind die Version von ZooKeeper aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. ZooKeeper

Informationen zur Version der Komponenten, mit denen ZooKeeper in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 6.15.0](#).

ZooKeeper Versionsinformationen für emr-6.15.0

| EMRAmazon-Freigabeetikett | ZooKeeper Ausführung | Komponenten, die mit installiert wurden ZooKeeper |
|---------------------------|----------------------|---|
| emr-6.15.0 | ZooKeeper 3.5.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

In der folgenden Tabelle sind die Version von ZooKeeper aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, sowie die Komponenten, mit denen Amazon EMR installiert. ZooKeeper

Informationen zur Version der Komponenten, mit denen ZooKeeper in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

ZooKeeper Versionsinformationen für emr-5.36.2

| EMRAmazon-Freigabeetikett | ZooKeeper Ausführung | Komponenten, die mit installiert wurden ZooKeeper |
|---------------------------|----------------------|---|
| emr-5.36.2 | ZooKeeper 3.4.14 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

Themen

- [ZooKeeper Verlauf der Veröffentlichung](#)

ZooKeeper Verlauf der Veröffentlichung

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon ZooKeeper enthalten sindEMR, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Release-Versionen von Amazon EMR 7.x](#), [Release-Versionen von Amazon EMR 6.x](#) oder [Amazon EMR 5.x-Release-Versionen](#).

ZooKeeper Versionsinformationen

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-6.11.1 | 3.5.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.11.0 | 3.5.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.10.1 | 3.5.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.35.0 | 3.4.14 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.5.0 | 3.5.7 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.4.0 | 3.5.7 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|---|
| emr-6.3.1 | 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-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-6.2.0 | 3.4.14 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.1.1 | 3.4.14 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.1.0 | 3.4.14 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.33.1 | 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 |
| emr-5.33.0 | 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 |
| emr-5.32.1 | 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|---|
| emr-5.30.2 | 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-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.25.0 | 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 |
| emr-5.24.1 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.24.0 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.21.2 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.21.1 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.21.0 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.20.1 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.20.0 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.19.1 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.19.0 | 3.4.13 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.18.1 | 3.4.12 | 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-5.18.0 | 3.4.12 | 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.16.1 | 3.4.12 | 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-5.16.0 | 3.4.12 | 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-5.15.1 | 3.4.12 | 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|---|
| emr-5.15.0 | 3.4.12 | 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, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.14.0 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.12.3 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.12.2 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.12.1 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.12.0 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|---|
| emr-5.11.2 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|---|
| emr-5.7.0 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.5.4 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.5.3 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.5.2 | 3.4.10 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|--|
| emr-5.4.0 | 3.4.9 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.3.2 | 3.4.9 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.3.1 | 3.4.9 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit ZooKeeper |
|----------------------------|----------------------|---|
| emr-5.3.0 | 3.4.9 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.2.3 | 3.4.9 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.2.2 | 3.4.9 | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, 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-resource-manager, zookeeper-client, zookeeper-server |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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 |

| Amazon EMR Release-Etikett | ZooKeeper Ausführung | Komponenten wurden installiert mit 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-resource-manager, 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-resource-manager, zookeeper-client, zookeeper-server |

Konnektoren und Dienstprogramme

Amazon EMR bietet mehrere Konnektoren und Dienstprogramme für den Zugriff auf andere AWS Dienste als Datenquellen. Sie können in der Regel auf Daten in diesen Services innerhalb eines Programms zugreifen. Sie können beispielsweise einen Kinesis-Stream in einer Hive-Abfrage, einem Pig-Skript oder einer MapReduce Anwendung angeben und dann mit diesen Daten arbeiten.

Themen

- [Exportieren, Importieren, Abfragen und Verbinden von Tabellen in DynamoDB mithilfe von Amazon EMR](#)
- [Kinesis](#)
- [S3 DistCp \(s3-dist-cp\)](#)
- [Aufräumen nach fehlgeschlagenen DistCp S3-Jobs](#)

Exportieren, Importieren, Abfragen und Verbinden von Tabellen in DynamoDB mithilfe von Amazon EMR

Note

Der Amazon EMR -DynamoDB Connector ist als Open Source verfügbar auf. GitHub Weitere Informationen finden Sie unter. <https://github.com/awslabs/emr-dynamodb-connector>


DynamoDB ist ein vollständig verwalteter Dienst ohne SQL Datenbank, der schnelle und vorhersehbare Leistung mit nahtloser Skalierbarkeit bietet. Entwickler können eine Datenbanktabelle erstellen und den durch Anfragen erzeugten Datenverkehr oder den Speicher unbegrenzt erweitern. DynamoDB verteilt die Daten und den Datenverkehr für die Tabelle automatisch auf eine ausreichende Anzahl von Servern, um die vom Kunden angegebene Anforderungskapazität zu erreichen und die gespeicherte Datenmenge zu verarbeiten und dabei gleichzeitig eine konsistente, schnelle Leistung beizubehalten. Mit Amazon EMR und Hive können Sie große Datenmengen, wie z. B. in DynamoDB gespeicherte Daten, schnell und effizient verarbeiten. Weitere Informationen über DynamoDB finden Sie im [Entwicklerhandbuch von Amazon DynamoDB](#).

Apache Hive ist eine Softwareschicht, mit der Sie Map-Reduce-Cluster mithilfe einer vereinfachten, SQL ähnlichen Abfragesprache namens HiveQL abfragen können. Die Software baut auf der

Hadoop-Architektur auf. Weitere Informationen zu Hive und HiveQL erhalten Sie in der [HiveQL-Sprachreferenz](#). Weitere Informationen zu Hive und Amazon finden Sie EMR unter [Apache Hive](#).

Sie können Amazon EMR mit einer angepassten Version von Hive verwenden, die Konnektivität zu DynamoDB bietet, um Operationen mit in DynamoDB gespeicherten Daten durchzuführen:

- DynamoDB-Daten in das Hadoop Distributed File System (HDFS) laden und als Eingabe in einen Amazon-Cluster verwenden. EMR
- Abfragen von DynamoDB-Live-Daten mit SQL ähnlichen Anweisungen (HiveQL).
- Verknüpfen von in DynamoDB gespeicherten Daten und deren Export oder Abfragen der verknüpften Daten.
- Exportieren von in DynamoDB gespeicherten Daten nach Amazon S3.
- Importieren von in Amazon S3 gespeicherten Daten in DynamoDB.

 Note

[Der Amazon EMR -DynamoDB-Connector unterstützt keine Cluster, die für die Verwendung der Kerberos-Authentifizierung konfiguriert sind.](#)

Um jede der folgenden Aufgaben auszuführen, starten Sie einen EMR Amazon-Cluster, geben den Speicherort der Daten in DynamoDB an und geben Hive-Befehle aus, um die Daten in DynamoDB zu bearbeiten.

Es gibt mehrere Möglichkeiten, einen EMR Amazon-Cluster zu starten: Sie können die EMR Amazon-Konsole, die Befehlszeilenschnittstelle (CLI) verwenden oder Ihren Cluster mit einem AWS SDK oder Amazon programmieren EMRAPI. Sie können auch bestimmen, ob ein Hive-Cluster interaktiv oder über ein Skript ausgeführt werden soll. In diesem Abschnitt zeigen wir Ihnen, wie Sie einen interaktiven Hive-Cluster von der EMR Amazon-Konsole aus starten und CLI

Die interaktive Verwendung von Hive ist eine hervorragende Möglichkeit zum Testen der Abfrageleistung und Optimieren Ihrer Anwendung. Nachdem Sie eine Reihe von Hive-Befehlen eingerichtet haben, die regelmäßig ausgeführt werden, sollten Sie erwägen, ein Hive-Skript zu erstellen, das Amazon für Sie ausführen EMR kann.

Warning

EMR-Lese- oder Schreibvorgänge von Amazon an einer DynamoDB-Tabelle werden auf Ihren festgelegten bereitgestellten Durchsatz angerechnet, wodurch sich möglicherweise die Häufigkeit von Ausnahmen beim bereitgestellten Durchsatz erhöht. Bei großen Anfragen implementiert Amazon Wiederholungen mit exponentiellem Backoff, um die Anforderungslast in der DynamoDB-Tabelle zu verwalten. Wenn Sie EMR Amazon-Jobs gleichzeitig mit anderem Datenverkehr ausführen, kann dies dazu führen, dass Sie den zugewiesenen, bereitgestellten Durchsatz überschreiten. Sie können dies überwachen, indem Sie die `ThrottleRequests`-Metrik in Amazon überprüfen CloudWatch. Wenn die Anforderungslast zu hoch ist, können Sie den Cluster neu starten und den Wert [Einstellung der Leserate in Prozent](#) oder [Einstellung der Schreibrate in Prozent](#) auf einen niedrigeren Wert setzen, um den EMR Amazon-Betrieb zu drosseln. Weitere Informationen zu DynamoDB-Durchsatzeinstellungen finden Sie unter [Bereitgestellter Durchsatz](#). Wenn eine Tabelle für den [On-Demand-Modus](#) konfiguriert ist, sollten Sie die Tabelle wieder in den Bereitstellungsmodus ändern, bevor Sie einen Export- oder Importvorgang ausführen. Pipelines benötigen ein Durchsatzverhältnis, um die zu verwendenden Ressourcen anhand eines D berechnen zu können. `ynamoDBtable` Im On-Demand-Modus wird der bereitgestellte Durchsatz entfernt. Um Durchsatzkapazität bereitzustellen, können Sie Amazon CloudWatch Events-Metriken verwenden, um den Gesamtdurchsatz auszuwerten, den eine Tabelle verwendet hat.

Themen

- [Eine Hive-Tabelle einrichten, um Hive-Befehle auszuführen](#)
- [Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten in DynamoDB](#)
- [Optimierung der Leistung für EMR Amazon-Operationen in DynamoDB](#)

Eine Hive-Tabelle einrichten, um Hive-Befehle auszuführen

Apache Hive ist eine Data Warehouse-Anwendung, mit der Sie in EMR Amazon-Clustern enthaltene Daten in einer SQL-ähnlichen Sprache abfragen können. Weitere Informationen zu Hive finden Sie unter <http://hive.apache.org/>.

Beim folgenden Verfahren wird davon ausgegangen, dass Sie bereits einen Cluster erstellt und ein EC2 Amazon-Schlüsselpaar angegeben haben. Informationen zu den ersten Schritten beim Erstellen von Clustern finden Sie unter [Erste Schritte mit Amazon EMR](#) im Amazon EMR Management Guide.

Konfigurieren Sie Hive für die Verwendung MapReduce

Wenn Sie Hive on Amazon verwenden, EMR um DynamoDB-Tabellen abzufragen, können Fehler auftreten, wenn Hive die Standard-Ausführungs-Engine Tez verwendet. Aus diesem Grund empfehlen wir, bei der Erstellung eines Clusters mit Hive, der wie in diesem Abschnitt beschrieben in DynamoDB integriert ist, eine Konfigurationsklassifizierung zu verwenden, die Hive auf die Verwendung festlegt. MapReduce Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Der folgende Ausschnitt zeigt die Konfigurationsklassifizierung und die Eigenschaft, die verwendet werden sollen, um sie MapReduce als Ausführungs-Engine für Hive festzulegen:

```
[
    {
        "Classification": "hive-site",
        "Properties": {
            "hive.execution.engine": "mr"
        }
    }
]
```

Interaktives Ausführen von Hive-Befehlen

1. Herstellen einer Verbindung mit dem Master-Knoten. Weitere Informationen finden Sie unter [Connect dem Master-Knoten herstellen SSH](#) im Amazon EMR Management Guide.
2. Geben Sie an der Eingabeaufforderung für den aktuellen Master-Knoten `hive` ein.

Sie sollten eine Hive-Eingabeaufforderung sehen: `hive>`

3. Geben Sie einen Hive-Befehl ein, der die Daten in DynamoDB einer Tabelle in der Hive-Anwendung zuordnet. Diese Tabelle fungiert als Referenz für die in Amazon DynamoDB gespeicherten Daten. Die Daten werden nicht lokal in Hive gespeichert und alle Abfragen, die diese Tabelle verwenden, werden mit den Live-Daten in DynamoDB abgeglichen. Dabei wird die Lese- und Schreibkapazität der Tabelle jedes Mal verbraucht, wenn ein Befehl ausgeführt wird. Wenn Sie mehrere Hive-Befehle für denselben Datensatz ausführen möchten, sollten Sie ihn zunächst exportieren.

Das folgende Beispiel zeigt die Syntax für die Zuweisung einer Hive-Tabelle zu einer DynamoDB-Tabelle.

```
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...")
```

Wenn Sie in Hive eine Tabelle mithilfe der DynamoDB erstellen möchten, müssen Sie sie unter Verwendung des Schlüsselworts EXTERNAL als externe Tabelle erstellen. Der Unterschied zwischen internen und externen Tabellen liegt darin, dass die Daten in internen Tabellen gelöscht werden, wenn die interne Tabelle gelöscht wird. Dies ist nicht das gewünschte Verhalten, wenn eine Verbindung zur Amazon DynamoDB besteht und somit werden nur externe Tabellen unterstützt.

Beispielsweise wird mit dem folgenden Befehl eine Tabelle mit dem Namen `hivetable1` in Hive erstellt, die auf die DynamoDB-Tabelle namens `dynamodbtable1` verweist. Die DynamoDB-Tabelle `dynamodbtable1` hat ein Primärschlüsselschema. `hash-and-range` Das Hash-Schlüsselement ist `name` (Zeichenfolgetyp), das Bereichsschlüsselement ist `year` (numerischer Typ) und jedes Element verfügt über einen Attributwert für `holidays` (Zeichenfolgetyp).

```
CREATE EXTERNAL TABLE hivetable1 (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");
```

Zeile 1 verwendet die HiveQL-Anweisung CREATE EXTERNAL TABLE. Für `hivetable1` müssen Sie für jedes Attributnamen-Wertpaar in der DynamoDB-Tabelle eine Spalte erstellen und den Datentyp angeben. Für diese Werte muss die Groß- und Kleinschreibung nicht beachtet werden und Sie können den Spalten beliebige Namen geben (außer reservierten Wörtern).

Zeile 2 verwendet die `STORED BY`-Anweisung. Bei dem Wert `STORED BY` handelt es sich um den Namen der Klasse, die die Verbindung zwischen Hive und DynamoDB handhabt. Er sollte auf `'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'` festgelegt werden.

Zeile 3 verwendet die `TBLPROPERTIES`-Anweisung, um „hivetable1“ mit der richtigen Tabelle und dem Schema in DynamoDB zu verknüpfen. Stellen Sie `TBLPROPERTIES` die Werte für den `dynamodb.table.name`-Parameter und `dynamodb.column.mapping`-Parameter bereit. Bei diesen Werten muss die Groß- und Kleinschreibung beachtet werden.

Note

Alle DynamoDB-Attributnamen für die Tabelle müssen über entsprechende Spalten in der Hive-Tabelle verfügen. Abhängig von Ihrer EMR Amazon-Version treten die folgenden Szenarien auf, wenn die one-to-one Zuordnung nicht existiert:

- In EMR Amazon-Version 5.27.0 und höher verfügt der Connector über Validierungen, die eine one-to-one Zuordnung zwischen DynamoDB-Attributnamen und Spalten in der Hive-Tabelle sicherstellen. Wenn die Zuordnung nicht existiert, tritt ein Fehler auf. one-to-one
- In EMR Amazon-Version 5.26.0 und früher enthält die Hive-Tabelle das Name-Wert-Paar aus DynamoDB nicht. Wenn Sie die Primärschlüsselattribute der DynamoDB nicht zuordnen, generiert Hive einen Fehler. Wenn Sie ein nicht-primäres Schlüsselattribut nicht zuordnen, wird kein Fehler generiert, aber die Daten werden auch nicht in der Hive-Tabelle angezeigt. Wenn die Datentypen nicht übereinstimmen, ist der Wert Null.

Anschließend können Sie die Hive-Vorgänge in `hivetable1` starten. Abfragen für `hivetable1` werden intern über die DynamoDB-Tabelle `dynamodhtable1` Ihres DynamoDB-Kontos ausgeführt, wodurch bei jeder Ausführung Lese- oder Schreibeinheiten verbraucht werden.

Wenn Sie Hive-Abfragen über eine DynamoDB-Tabelle ausführen, müssen Sie sicherstellen, dass ausreichend Lesekapazitätseinheiten vorhanden sind.

Nehmen Sie zum Beispiel an, dass Sie 100 Lesekapazitätseinheiten für Ihre DynamoDB-Tabelle bereitgestellt haben. Dadurch können Sie 100 Lesevorgänge oder 409.600 Byte pro Sekunde

ausführen. Wenn diese Tabelle 20 GB an Daten enthält (21.474.836.480 Bytes) und Ihre Hive-Abfrage einen vollständigen Tabellen-Scan durchführt, können Sie ermitteln, wie lange die Ausführung der Abfrage dauern wird:

$$21\,474\,836\,480 / 409\,600 = 52\,429 \text{ Sekunden} = 14,56 \text{ Stunden}$$

Die einzige Möglichkeit, um die erforderliche Zeit zu verkürzen, würde eine Änderung der Lesekapazitätseinheiten in der DynamoDB-Quelltabelle notwendig machen. Das Hinzufügen weiterer EMR Amazon-Knoten hilft nicht.

Der Fertigstellungsgrad in der Hive-Ausgabe wird dann aktualisiert, wenn ein oder mehrere Mapper-Prozesse abgeschlossen wurden. Für eine große DynamoDB-Tabelle mit geringer Lesekapazität wird der Fertigstellungsgrad der Ausgabe ggf. für längere Zeit nicht aktualisiert. Im oben genannten Fall wird der Auftrag mehrere Stunden lang als zu 0 % abgeschlossen angezeigt. Genauere Informationen zum Status Ihres Jobs finden Sie in der EMR Amazon-Konsole. Dort können Sie den Status der einzelnen Mapper-Aufgaben sowie Statistiken zu den gelesenen Daten einsehen. Sie können sich auch bei der Hadoop-Schnittstelle auf dem Master-Knoten anmelden und die Hadoop-Statistiken anzeigen. Hier werden Ihnen der Status einzelner Map-Aufgaben sowie Statistiken für Lesevorgänge angezeigt. Weitere Informationen finden Sie unter den folgenden Themen:

- [Auf Hauptknoten gehostete Web-Schnittstellen](#)
- [Hadoop-Web-Schnittstellen anzeigen](#)

Weitere Informationen über HiveQL-Beispielanweisungen, um Aufgaben auszuführen, die bei der Ausführung von Aufgaben, wie dem Exportieren oder Importieren von DynamoDB-Daten sowie dem Verknüpfen von Tabellen behilflich sind, finden Sie unter [Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten in DynamoDB](#).

Abbrechen einer Hive-Anfrage

Wenn Sie eine Hive-Abfrage ausführen, enthält die Erstantwort vom Server den Befehl, die Anfrage abubrechen. Um die Anfrage zu jedem Zeitpunkt während des Vorgangs abubrechen, verwenden Sie den Kill Command (Beendigungsbefehl) aus der Serverantwort.

1. Geben Sie `Ctrl+C` ein, um den Befehlszeilen-Client zu beenden.
2. Geben Sie nach der Shell-Eingabeaufforderung den Kill Command (Beendigungsbefehl) aus der Erstantwort des Servers auf Ihre Anfrage ein.

Alternativ können Sie den folgenden Befehl von der Befehlszeile des Master-Knotens aus ausführen, um den Hadoop-Job zu beenden. Dabei *job-id* ist der Bezeichner des Hadoop-Jobs und kann von der Hadoop-Benutzeroberfläche abgerufen werden.

```
hadoop job -kill job-id
```

Datentypen für Hive und DynamoDB

Die folgende Tabelle zeigt die verfügbaren Hive-Datentypen, den DynamoDB-Standardtyp, dem sie entsprechen, und die alternativen DynamoDB-Typen, denen sie auch zugeordnet werden können.

| Hive-Typ | Standardtyp DynamoDB | Alternative DynamoDB-Typ |
|---------------------|----------------------|---|
| Zeichenfolge | String (S) | |
| Bigint oder Doppel | Ganzzahl | |
| Binary | Binary (B) | |
| boolesch | boolean () BOOL | |
| Array | list (L) | Zahlensatz (ZS), ein Zeichenfolgesatz (ZFS) oder ein Binärsatz (BS) |
| map<string, string> | Element | map (M) |
| map<string, ?> | map (M) | |
| | null () NULL | |

Wenn Sie Ihre Hive-Daten als entsprechenden alternativen DynamoDB-Typ schreiben möchten oder wenn Ihre DynamoDB-Daten Attributwerte eines alternativen -Typs enthalten, können Sie die Spalte

und den DynamoDB-Typ mit dem Parameter `dynamodb.type.mapping` angeben. Das folgende Beispiel zeigt die Syntax für die Angabe einer alternativen Typzuweisung.

```
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");
```

Der Parameter für die Typzuweisung ist optional und muss nur für die Spalten angegeben werden, die alternative Typen verwenden.

Beispielsweise erstellt der folgende Hive-Befehl eine Tabelle mit dem Namen `hivetable2`, die auf die DynamoDB-Tabelle `dynamodbttable2` verweist. Sie ist mit `hivetable1` vergleichbar, außer dass sie die `col3`-Spalte dem String Set(SS)-Typ zuordnet.

```
CREATE EXTERNAL TABLE hivetable2 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbttable2",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays",
"dynamodb.type.mapping" = "col3:SS");
```

In Hive sind `hivetable1` und `hivetable2` identisch. Wenn jedoch Daten aus diesen Tabellen in die entsprechenden DynamoDB-Tabellen geschrieben werden, enthält `dynamodbttable1` Listen, während `dynamodbttable2` Zeichenfolgensätze enthält.

Wenn Sie Hive-null-Werte als Attribute des DynamoDB-null-Typs schreiben möchten, ist dies mit dem Parameter `dynamodb.null.serialization` möglich. Das folgende Beispiel zeigt die Syntax für die Angabe der null-Serialisierung.

```
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");
```

Der Serialisierungsparameter `Null` ist optional und wird auf `false` gesetzt, wenn er nicht angegeben wird. Beachten Sie, dass DynamoDB `null`-Attribute unabhängig von der Parametereinstellung als `null`-Werte in Hive gelesen werden. Hive-Sammlungen mit `null`-Werten können nur dann in DynamoDB geschrieben werden, wenn der `Null`-Serialisierungsparameter als `true` angegeben ist. Andernfalls tritt ein Hive-Fehler auf.

Der `bigint`-Typ in Hive ist identisch mit dem Java-Typ "long" und der Hive-Doppeltyp entspricht mit dem Java-Doppeltyp im Hinblick auf die Genauigkeit. Das bedeutet, dass Sie für Ihre in DynamoDB gespeicherten numerischen Daten, deren Präzision höher liegt als in den Hive-Datentypen, bei der Verwendung von Hive für den Export oder Import oder bei einem Verweis auf die DynamoDB-Daten mit einem Präzisionsverlust oder einem Scheitern der Hive-Abfrage rechnen müssen.

Exporte des binären Typs von DynamoDB nach Amazon Simple Storage Service (Amazon S3) oder HDFS werden als Base64-kodierte Zeichenfolge gespeichert. Wenn Sie Daten aus Amazon S3 oder HDFS in den DynamoDB-Binärtyp importieren, sollten sie als Base64-Zeichenfolge codiert werden.

Hive-Optionen

Sie können die folgenden Hive-Optionen festlegen, um die Datenübertragung aus Amazon DynamoDB zu verwalten. Diese Optionen gelten nur für die aktuelle Hive-Sitzung. Wenn Sie die Hive-Eingabeaufforderung schließen und später auf dem Cluster erneut öffnen, werden diese Einstellungen auf die Standardwerte zurückgesetzt.

| Hive-Optionen | Beschreibung |
|---|--|
| <code>dynamodb.throughput.read.percent</code> | <p>Legt die Rate der Lesevorgänge so fest, dass Ihre von DynamoDB bereitgestellte Durchsatzrate im für Ihre Tabelle zugewiesenen Bereich liegt. Der Wert liegt zwischen 0.1 und 1.5 (einschließlich).</p> <p>Der Wert für die Standardleserate liegt bei 0,5. Das bedeutet, dass Hive versuchen wird, die Hälfte der Lesekapazität, die für Ressourcen in der Tabelle bereitgestellt ist, zu verbrauchen. Eine Erhöhung dieses Werts auf über 0,5 verbessert die Leseanforderungsrate. Eine Verringerung dieses Werts unter 0,5 verringert die Leseanforderungsrate. Bei dieser Leserate handelt es sich um einen ungefähren</p> |

| Hive-Optionen | Beschreibung |
|---------------|--|
| | <p>Wert. Die tatsächliche Leserate variiert abhängig von Faktoren wie der Tatsache, ob eine einheitliche Verteilung der Schlüssel in DynamoDB vorliegt.</p> <p>Wenn Sie feststellen, dass der bereitgestellte Durchsatz häufig durch die Hive-Operation überschritten wird oder wenn der Live-Lesedatenverkehr zu sehr gedrosselt wird, reduzieren Sie diesen Wert auf unter 0,5. Wenn Sie über genügend Kapazität verfügen und sich schnellere Hive-Operationen wünschen, legen Sie den Wert auf über 0,5 fest. Sie können auch überzeichnen, indem Sie den Wert auf 1,5 festlegen, wenn Sie der Ansicht sind, dass ungenutzte Ein-/Ausgabevorgänge verfügbar sind.</p> |

| Hive-Optionen | Beschreibung |
|--|---|
| <code>dynamodb.throughput.write.percent</code> | <p>Legt die Rate der Schreibvorgänge so fest, dass Ihre von DynamoDB bereitgestellte Durchsatzrate im für Ihre Tabelle zugewiesenen Bereich liegt. Der Wert liegt zwischen 0.1 und 1.5 (einschließlich).</p> <p>Der Wert für die Standardschreibrate liegt bei 0,5. Das bedeutet, dass Hive versuchen wird, die Hälfte der Schreibkapazität, die für Ressourcen in der Tabelle bereitgestellt ist, zu verbrauchen. Eine Erhöhung dieses Werts auf über 0,5 verbessert die Schreibankorderungsrate. Eine Verringerung dieses Werts unter 0,5 verringert die Schreibankorderungsrate. Bei dieser Schreibrate handelt es sich um einen ungefähren Wert. Die tatsächliche Schreibrate variiert abhängig von Faktoren wie der Tatsache, ob eine einheitliche Verteilung der Schlüssel in DynamoDB vorliegt.</p> <p>Wenn Sie feststellen, dass der bereitgestellte Durchsatz häufig durch die Hive-Operation überschritten wird oder wenn der Live-Schreibdatenverkehr zu sehr gedrosselt wird, reduzieren Sie diesen Wert auf unter 0.5. Wenn Sie über genügend Kapazität verfügen und sich schnellere Hive-Operationen wünschen, legen Sie den Wert auf über 0.5 fest. Sie können auch überzeichnen, indem Sie den Wert auf 1,5 festlegen, wenn Sie der Ansicht sind, dass ungenutzte Ein-/Ausgabevorgänge verfügbar sind oder es sich hierbei um das anfängliche Hochladen von Daten in die Tabelle handelt und noch kein aktiver Datenverkehr vorliegt.</p> |
| <code>dynamodb.endpoint</code> | Geben Sie den Endpunkt für den DynamoDB-Service an. Weitere Informationen über die verfügbaren DynamoDB-Endpunkte finden Sie unter Regionen und Endpunkte . |

| Hive-Optionen | Beschreibung |
|--------------------------------------|---|
| <code>dynamodb.max.map.tasks</code> | Geben Sie die maximale Anzahl von Zuordnungsaufgaben beim Lesen von Daten aus DynamoDB an. Dieser Wert muss gleich oder größer 1 sein. |
| <code>dynamodb.retry.duration</code> | Geben Sie die Anzahl der Minuten an, die als Timeout-Dauer für das Wiederholen der Hive-Befehle verwendet werden soll. Bei diesem Wert muss es sich um eine Ganzzahl gleich oder größer als 0 handeln. Die Standard-Timeout-Dauer beträgt zwei Minuten. |

Diese Optionen werden über den Befehl SET, wie im folgenden Beispiel gezeigt, festgelegt.

```
SET dynamodb.throughput.read.percent=1.0;  
  
INSERT OVERWRITE TABLE s3_export SELECT *  
FROM hiveTableName;
```

Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten in DynamoDB

In den folgenden Beispielen werden Hive-Befehle verwendet, um Operationen wie das Exportieren von Daten nach Amazon S3 oder das Importieren von Daten in DynamoDBHDFS, das Verbinden von Tabellen, das Abfragen von Tabellen und mehr auszuführen.

Operationen für eine Hive-Tabelle verweisen auf Daten, die in DynamoDB gespeichert sind. Hive-Befehle sind abhängig von den Einstellungen der DynamoDB-Tabelle für den bereitgestellten Durchsatz. Die Daten, die abgerufen werden, umfassen die Daten, die zum Zeitpunkt der Verarbeitung der Hive-Operationsanforderung durch DynamoDB in die DynamoDB-Tabelle geschrieben werden. Wenn der Datenabruf viel Zeit in Anspruch nimmt, wurden einige der vom Hive-Befehl zurückgegebenen Daten möglicherweise seit Starten des Hive-Befehls in DynamoDB aktualisiert.

Die Hive-Befehle `DROP TABLE` und `CREATE TABLE` gelten nur für die lokalen Tabellen in Hive und erstellen bzw. löschen keine Tabellen in DynamoDB. Wenn Ihre Hive-Abfrage auf eine Tabelle in DynamoDB verweist, muss diese bereits vorhanden sein, bevor Sie die Abfrage ausführen. Weitere Informationen zum Erstellen und Löschen von Tabellen in DynamoDB finden Sie unter [Arbeiten mit Tabellen in DynamoDB](#) im Amazon-DynamoDB-Entwicklerhandbuch.

Note

Wenn Sie eine Hive-Tabelle mit einem Speicherort in Amazon S3 verknüpfen, weisen Sie sie nicht dem Stammpfad des Buckets `s3://mybucket` zu, da dies Fehler verursachen kann, wenn Hive die Daten nach Amazon S3 schreibt. Weisen Sie die Tabelle stattdessen einem Unterpfad des Buckets zu ("`s3://mybucket/mypath`").

Exportieren von Daten aus DynamoDB

Sie können mit Hive Daten aus DynamoDB exportieren.

So exportieren Sie eine DynamoDB-Tabelle in einen Amazon-S3-Bucket

- Erstellen Sie eine Hive-Tabelle, die auf in DynamoDB gespeicherte Daten verweist. Anschließend können Sie den `INSERT OVERWRITE` Befehl aufrufen, um die Daten in ein externes Verzeichnis zu schreiben. Im folgenden Beispiel `s3://bucketname/path/subpath/` ist ein gültiger Pfad in Amazon S3. Passen Sie die Spalten und Datentypen im `CREATE` Befehl an die Werte in Ihrer DynamoDB an. Sie können diese Methode zum Erstellen eines Archivs Ihrer DynamoDB-Daten in Amazon S3 verwenden.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name, col2:year, col3:holidays");

INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *
FROM hiveTableName;
```

So exportieren Sie eine DynamoDB-Tabelle in einen Amazon S3-Bucket mit Formatierung

- Erstellen Sie eine externe Tabelle, die auf einen Speicherort in Amazon S3 verweist. Dies wird unten als "s3_export" gezeigt. Geben Sie während des CREATE Aufrufs die Zeilenformatierung für die Tabelle an. Wenn Sie dann Daten von DynamoDB INSERT OVERWRITE nach s3_export exportieren, werden die Daten im angegebenen Format ausgegeben. Im folgenden Beispiel werden die Daten als kommagetrennte Werte (,) ausgegeben. CSV

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodatabase1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

So exportieren Sie eine DynamoDB-Tabelle in einen Amazon-S3-Bucket ohne Angabe einer Spaltenzuordnung

- Erstellen Sie eine Hive-Tabelle, die auf in DynamoDB gespeicherte Daten verweist. Dies ähnelt dem vorhergehenden Beispiel mit der Ausnahme, dass Sie keine Spaltenzuordnung angeben. Die Tabelle muss über genau einen Spaltentyp `map<string, string>` verfügen. Wenn Sie dann eine EXTERNAL-Tabelle in Amazon S3 erstellen, können Sie den Befehl INSERT OVERWRITE zum Schreiben der Daten von DynamoDB nach Amazon S3 aufrufen. Sie können diese Methode zum Erstellen eines Archivs Ihrer DynamoDB-Daten in Amazon S3 verwenden. Da es keine Spaltenzuordnung gibt, können Sie keine Tabellen abfragen, die auf diese Weise exportiert werden. Das Exportieren von Daten ohne Angabe einer Spaltenzuordnung ist in Hive 0.8.1.5 oder höher verfügbar, das auf Amazon 2.2 unterstützt wird. EMR AMI x und höher.

```
CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
```

```
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbt1");

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;
```

So exportieren Sie eine DynamoDB-Tabelle in einen Amazon-S3-Bucket mit Datenkomprimierung

- Hive bietet mehrere Kompressions-Codecs an, die Sie während der Hive-Sitzung einrichten können. Dadurch werden die exportierten Daten im angegebenen Format komprimiert. Im folgenden Beispiel werden die exportierten Dateien mit dem Lempel-Ziv-Oberhumer () - Algorithmus komprimiert. 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" = "dynamodbt1",
"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;
```

Die verfügbaren Kompressions-Codecs sind:

- 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`

So exportieren Sie eine DynamoDB-Tabelle nach HDFS

- Verwenden Sie den folgenden Hive-Befehl, wobei `hdfs:///directoryName` ist ein gültiger HDFS Pfad und `hiveTableName` ist eine Tabelle in Hive, die auf DynamoDB verweist. Dieser Exportvorgang ist schneller als der Export einer DynamoDB-Tabelle nach Amazon S3, da Hive 0.7.1.1 ihn HDFS als Zwischenschritt beim Exportieren von Daten nach Amazon S3 verwendet. Das folgende Beispiel zeigt, wie Sie `dynamodb.throughput.read.percent` auf 1.0 festlegen, um die Leseanforderungsrate zu erhöhen.

```
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;
```

Sie können Daten auch exportieren, indem Sie die Formatierung und Komprimierung HDFS verwenden, wie oben für den Export nach Amazon S3 gezeigt. Ersetzen Sie dazu einfach das Amazon S3 S3-Verzeichnis in den obigen Beispielen durch ein HDFS Verzeichnis.

Um nicht druckbare Daten mit UTF 8 Zeichen in Hive zu lesen

- Sie können mit Hive nicht druckbare Daten mit einer Länge von UTF 8 Zeichen lesen und schreiben, indem Sie die `STORED AS SEQUENCEFILE` Klausel verwenden, wenn Sie die Tabelle erstellen. A SequenceFile ist das Hadoop-Binärdateiformat. Sie müssen Hadoop verwenden, um diese Datei zu lesen. Das folgende Beispiel zeigt, wie Daten aus DynamoDB nach Amazon S3 exportiert werden. Sie können diese Funktion verwenden, um nicht druckbare UTF -8-kodierte Zeichen zu verarbeiten.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtbl1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE 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;
```

Importieren von Daten in DynamoDB

Wenn Sie mit Hive Daten in DynamoDB schreiben möchten, stellen Sie sicher, dass die Anzahl der Schreibkapazitätseinheiten größer als die Anzahl der Mapper im Cluster ist. Cluster, die auf EC2 m1.xlarge-Instances ausgeführt werden, erzeugen beispielsweise 8 Mapper pro Instanz. Bei einem Cluster mit 10 Instances ergäbe dies 80 Mapper. Wenn Ihre Schreibkapazitätseinheiten nicht größer als die Anzahl der Mapper im Cluster sind, verbraucht der Hive-Schreibvorgang möglicherweise den gesamten Schreibdurchsatz oder versucht, mehr Durchsatz zu verbrauchen, als zur Verfügung gestellt wird. Weitere Informationen zur Anzahl der Mapper, die von jedem EC2 Instance-Typ erzeugt werden, finden Sie unter [Konfigurieren von Hadoop](#)

Die Anzahl der Mapper in Hadoop wird durch die Input Splits gesteuert. Wenn es zu wenig Splits gibt, ist Ihr Schreibbefehl ggf. nicht in der Lage, den gesamten verfügbaren Schreibdurchsatz zu verbrauchen.

Wenn ein Element mit demselben Schlüssel in der DynamoDB-Zieltabelle vorhanden ist, wird es überschrieben. Wenn kein Element mit dem Schlüssel in der DynamoDB-Zieltabelle vorhanden ist, wird das Element eingefügt.

So importieren Sie Daten aus Amazon S3 zu DynamoDB

- Sie können Amazon EMR (AmazonEMR) und Hive verwenden, um Daten von Amazon S3 nach DynamoDB zu schreiben.

```
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" = "dynamodatabase1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM s3_import;
```

So importieren Sie eine Tabelle von einem Amazon-S3-Bucket in DynamoDB ohne Angabe einer Spaltenzuordnung

- Erstellen Sie eine EXTERNAL-Tabelle, die auf die in Amazon S3 gespeicherten Daten verweist, die zuvor von DynamoDB exportiert wurden. Stellen Sie vor dem Import sicher, dass die Tabelle in DynamoDB vorhanden ist und über das gleiche Schlüsselschema wie die zuvor exportierte DynamoDB-Tabelle verfügt. Die Tabelle muss außerdem über genau einen Spaltentyp `map<string, string>` verfügen. Wenn Sie dann eine Hive-Tabelle erstellen, die mit DynamoDB verknüpft ist, können Sie den Befehl `INSERT OVERWRITE` zum Schreiben der Daten von Amazon S3 nach DynamoDB aufrufen. Da es keine Spaltenzuordnung gibt, können Sie keine Tabellen abfragen, die auf diese Weise importiert werden. Das Importieren von Daten ohne Angabe einer Spaltenzuordnung ist in Hive 0.8.1.5 oder höher verfügbar, was auf Amazon EMR AMI 2.2.3 und höher unterstützt wird.

```
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" = "dynamodatabase1");

INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

Um eine Tabelle von nach DynamoDB HDFS zu importieren

- Sie können Amazon EMR und Hive verwenden, um Daten von nach DynamoDB HDFS zu schreiben.

```
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;
```

Abfragen von Daten in DynamoDB

Die folgenden Beispiele zeigen die verschiedenen Möglichkeiten, wie Sie Amazon verwenden können EMR, um in DynamoDB gespeicherte Daten abzufragen.

So finden Sie den größten Wert für eine zugeordnete Spalte (**max**)

- Verwenden Sie Hive-Befehle wie die folgenden. Im ersten Befehl erstellt die CREATE Anweisung eine Hive-Tabelle, die auf in DynamoDB gespeicherte Daten verweist. Die SELECT Anweisung verwendet dann diese Tabelle, um in DynamoDB gespeicherte Daten abzufragen. Im folgenden Beispiel wird nach der größten Bestellung eines bestimmten Kunden gesucht.

```
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;
```

So aggregieren Sie Daten mit der **GROUP BY**-Klausel

- Sie können die GROUP BY-Klausel zum Sammeln von Daten über mehrere Datensätze hinweg verwenden. Diese Klausel wird häufig in Verbindung mit einer Aggregationsfunktion wie „sum“, „count“, „min“ oder „max“ eingesetzt. Das folgende Beispiel gibt eine Liste der größten Bestellungen von Kunden zurück, die mehr als drei Bestellungen aufgegeben haben.

```
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;
```

So verknüpfen Sie zwei DynamoDB-Tabellen

- Im folgenden Beispiel werden zwei Hive-Tabellen Daten in DynamoDB zugeordnet. Dann wird ein Join zwischen diesen beiden Tabellen aufgerufen. Der Join wird auf dem Cluster verarbeitet und zurückgegeben. Der Join wird nicht in DynamoDB ausgeführt. In diesem Beispiel wird eine Liste von Kunden mit ihren Einkäufen angezeigt, die mehr als zwei Bestellungen getätigt haben.

```
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;
```

So verknüpfen Sie zwei Tabellen aus verschiedenen Quellen

- Im folgenden Beispiel ist Customer_S3 eine Hive-Tabelle, die eine in Amazon S3 gespeicherte CSV Datei lädt, und hive_purchases ist eine Tabelle, die auf Daten in DynamoDB verweist. Das folgende Beispiel verknüpft Kundendaten, die als CSV Datei in Amazon S3 gespeichert sind, mit in DynamoDB gespeicherten Bestelldaten, um einen Datensatz zurückzugeben, der Bestellungen darstellt, die von Kunden aufgegeben wurden, deren Namen „Miller“ enthalten.

```
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

In den vorherigen Beispielen wurden die CREATE TABLE Aussagen aus Gründen der Übersichtlichkeit und Vollständigkeit in jedes Beispiel aufgenommen. Bei Ausführung mehrerer Abfragen oder Exportoperationen für eine bestimmte Hive-Tabelle müssen Sie die Tabelle nur einmal zu Beginn der Hive-Sitzung erstellen.

Optimierung der Leistung für EMR Amazon-Operationen in DynamoDB

EMRAmazon-Operationen an einer DynamoDB-Tabelle zählen als Lesevorgänge und unterliegen den bereitgestellten Durchsatzeinstellungen der Tabelle. Amazon EMR implementiert seine eigene Logik, um zu versuchen, die Last auf Ihrer DynamoDB-Tabelle auszugleichen, um die Möglichkeit zu minimieren, dass Ihr bereitgestellter Durchsatz überschritten wird. Am Ende jeder Hive-Abfrage EMR gibt Amazon Informationen über den Cluster zurück, der zur Verarbeitung der Abfrage verwendet wurde, einschließlich der Häufigkeit, mit der Ihr bereitgestellter Durchsatz überschritten wurde. Sie können diese Informationen sowie CloudWatch Metriken zu Ihrem DynamoDB-Durchsatz verwenden, um die Belastung Ihrer DynamoDB-Tabelle bei nachfolgenden Anfragen besser zu verwalten.

Die folgenden Faktoren beeinflussen die Hive-Abfrageleistung bei der Arbeit mit DynamoDB-Tabellen.

Bereitgestellte Lesekapazitätseinheiten

Wenn Sie Hive-Abfragen über eine DynamoDB-Tabelle ausführen, müssen Sie sicherstellen, dass ausreichend Lesekapazitätseinheiten vorhanden sind.

Nehmen Sie zum Beispiel an, dass Sie 100 Lesekapazitätseinheiten für Ihre DynamoDB-Tabelle bereitgestellt haben. Dadurch können Sie 100 Lesevorgänge oder 409.600 Byte pro Sekunde ausführen. Wenn diese Tabelle 20 GB an Daten enthält (21.474.836.480 Bytes) und Ihre Hive-Abfrage einen vollständigen Tabellen-Scan durchführt, können Sie ermitteln, wie lange die Ausführung der Abfrage dauern wird:

$$21\,474\,836\,480 / 409\,600 = 52\,429 \text{ Sekunden} = 14,56 \text{ Stunden}$$

Die einzige Möglichkeit, um die erforderliche Zeit zu verkürzen, würde eine Änderung der Lesekapazitätseinheiten in der DynamoDB-Quellentabelle notwendig machen. Das Hinzufügen weiterer Knoten zum EMR Amazon-Cluster hilft nicht.

Der Fertigstellungsgrad in der Hive-Ausgabe wird dann aktualisiert, wenn ein oder mehrere Mapper-Prozesse abgeschlossen wurden. Für eine große DynamoDB-Tabelle mit wenig bereitgestellter Lesekapazität wird die fertiggestellte Ausgabe in Prozent ggf. für längere Zeit nicht aktualisiert; im oben genannten Fall würde der Auftrag mehrere Stunden lang zu 0 % abgeschlossen angezeigt. Genauere Informationen zum Status Ihres Jobs finden Sie in der EMR Amazon-Konsole. Dort können Sie den Status der einzelnen Mapper-Aufgaben sowie Statistiken zu den gelesenen Daten einsehen.

Sie können sich auch bei der Hadoop-Schnittstelle auf dem Master-Knoten anmelden und die Hadoop-Statistiken anzeigen. Hier werden Ihnen der Status einzelner Map-Aufgaben sowie Statistiken für Lesevorgänge angezeigt. Weitere Informationen finden Sie [im Amazon EMR Management Guide unter Webschnittstellen, die auf dem Master-Knoten gehostet](#) werden.

Einstellung der Leserate in Prozent

Standardmäßig EMR verwaltet Amazon die Anforderungslast für Ihre DynamoDB-Tabelle entsprechend Ihrem aktuell bereitgestellten Durchsatz. Wenn Amazon jedoch Informationen über Ihren Job EMR zurückgibt, die eine große Anzahl von Antworten mit bereitgestelltem Durchsatz und Überschreitung enthalten, können Sie die Standardleserate mithilfe des `dynamodb.throughput.read.percent` Parameters anpassen, wenn Sie die Hive-Tabelle einrichten. Weitere Informationen zum Festlegen des Parameters zur Leserate in Prozent finden Sie unter [Hive-Optionen](#).

Einstellung der Schreibrate in Prozent

Standardmäßig EMR verwaltet Amazon die Anforderungslast für Ihre DynamoDB-Tabelle entsprechend Ihrem aktuell bereitgestellten Durchsatz. Wenn Amazon jedoch Informationen über Ihren Job EMR zurückgibt, die eine große Anzahl von Antworten mit bereitgestelltem Durchsatz und Überschreitung enthalten, können Sie die Standard-Schreibrate mithilfe des `dynamodb.throughput.write.percent` Parameters anpassen, wenn Sie die Hive-Tabelle einrichten. Weitere Informationen zum Festlegen des Parameters zur Schreibrate in Prozent finden Sie unter [Hive-Optionen](#).

Einstellung für das Wiederholungsintervall

Standardmäßig EMR führt Amazon eine Hive-Abfrage erneut aus, wenn innerhalb von zwei Minuten kein Ergebnis zurückgegeben wurde. Dies ist das standardmäßige Wiederholungsintervall. Sie können dieses Intervall anpassen, indem Sie bei der Ausführung einer Hive-Abfrage den Parameter `dynamodb.retry.duration` einstellen. Weitere Informationen zum Festlegen des Parameters zur Schreibrate in Prozent finden Sie unter [Hive-Optionen](#).

Anzahl der Zuordnungs-Tasks

Die Zuordnungs-Daemons, die Hadoop startet, um Ihre Anforderungen zum Exportieren und Abfragen von in DynamoDB gespeicherten Daten zu verarbeiten, sind auf eine maximale Leserate von 1 MiB pro Sekunde begrenzt, um die verbrauchte Lesekapazität einzuschränken. Wenn zusätzlich bereitgestellter Durchsatz auf DynamoDB vorhanden ist, verbessern Sie die Performance von Hive-Export- und Abfrageoperationen, indem Sie die Anzahl der Zuordnungs-Daemons erhöhen. Zu diesem Zweck können Sie entweder die Anzahl der EC2 Instances in Ihrem Cluster oder die Anzahl der Mapper-Daemons erhöhen, die auf jeder Instance ausgeführt werden. EC2

Sie können die Anzahl der EC2 Instanzen in einem Cluster erhöhen, indem Sie den aktuellen Cluster stoppen und ihn mit einer größeren Anzahl von Instanzen neu starten. EC2 Sie geben die Anzahl der EC2 Instances im Dialogfeld Configure EC2 Instances an, wenn Sie den Cluster von der EMR Amazon-Konsole aus starten, oder mit der `--num-instances` Option, wenn Sie den Cluster von der `aws emr` CLI aus starten.

Die Anzahl der Map-Tasks, die auf einer Instance ausgeführt werden, hängt vom EC2 Instance-Typ ab. Weitere Informationen zu den unterstützten EC2 Instance-Typen und der Anzahl der Mapper, die jeder Instance-Typ bereitstellt, finden Sie unter [Aufgabenkonfiguration](#). Hier finden Sie einen Abschnitt zur Aufgabenkonfiguration ("Task Configuration") für jede der unterstützten Konfigurationen.

Eine andere Möglichkeit, die Anzahl der Mapper-Daemons zu erhöhen, besteht darin, den Hadoop-Konfigurationsparameter `mapreduce.tasktracker.map.tasks.maximum` in einen höheren Wert zu ändern. Dies hat den Vorteil, dass Sie mehr Mapper erhalten, ohne die Anzahl oder Größe der EC2 Instanzen zu erhöhen, wodurch Sie Geld sparen. Ein Nachteil ist, dass ein zu hoher Wert dazu führen kann, dass den EC2 Instanzen in Ihrem Cluster der Arbeitsspeicher ausgeht. Zum Festlegen von `mapreduce.tasktracker.map.tasks.maximum` starten Sie den Cluster und geben einen Wert für `mapreduce.tasktracker.map.tasks.maximum` als Eigenschaft der Konfigurationsklassifizierung der `mapred-site` an. Dies wird im folgenden Beispiel veranschaulicht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
{
  "configurations": [
    {
      "classification": "mapred-site",
      "properties": {
        "mapred.tasktracker.map.tasks.maximum": "10"
      }
    }
  ]
}
```

```
]
}
```

Parallele Datenanforderungen

Mehrere Datenanforderungen, entweder von mehr als einem Benutzer oder von mehr als einer Anwendung, für eine einzelne Tabelle kann den bereitgestellten Lesedurchsatz erschöpfen und die Performance beeinträchtigen.

Prozessdauer

Die Datenkonsistenz in DynamoDB hängt von der Reihenfolge der Lese- und Schreibvorgänge auf den einzelnen Knoten ab. Während eine Hive-Abfrage verarbeitet wird, kann eine andere Anwendung neue Daten in die DynamoDB-Tabelle laden oder vorhandene Daten ändern oder löschen. In diesem Fall enthalten die Ergebnisse der Hive-Abfrage möglicherweise nicht die Datenänderungen, die vorgenommen wurden, während die Abfrage ausgeführt wurde.

Durchsatzüberschreitungen vermeiden

Wenn Sie Hive-Abfragen für DynamoDB ausführen, achten Sie darauf, dass Ihr bereitgestellter Durchsatz nicht überschritten wird, da sich dies nachteilig auf die Kapazität auswirkt, die für die DynamoDB: :Get-Aufrufe Ihrer Anwendung benötigt werden. Um sicherzustellen, dass dies nicht der Fall ist, sollten Sie das Lesevolumen und die Drosselung bei Anwendungsaufrufen regelmäßig überwachen, DynamoDB: :Get indem Sie die Protokolle und Messwerte in Amazon überprüfen. CloudWatch

Abfragezeit

Wenn Hive-Abfragen, die auf eine DynamoDB-Tabelle zugreifen, für Zeiten geplant werden, in denen wenig Anforderungen an die DynamoDB-Tabelle gerichtet werden, verbessert das die Leistung. Beispiel: Wenn die Mehrzahl der Anwendungsbenutzer in San Francisco lebt, können Sie die täglichen Daten um 04:00 Uhr PST, wenn die Mehrheit der Benutzer im Ruhezustand ist und keine Datensätze in Ihrer DynamoDB-Datenbank aktualisiert.

Zeitbasierte Tabellen

Wenn die Daten als eine Reihe von zeitbasierten DynamoDB-Tabellen organisiert sind, z. B. eine Tabelle pro Tag, können Sie die Daten exportieren, wenn die Tabelle nicht mehr aktiv ist. Sie können auf diese Weise Daten fortlaufend nach Amazon S3 sichern.

Archivierte Daten

Wenn Sie viele Hive-Abfragen für die in DynamoDB gespeicherten Daten ausführen möchten und Ihre Anwendung archivierte Daten tolerieren kann, sollten Sie die Daten nach Amazon S3 HDFS oder Amazon S3 exportieren und die Hive-Abfragen für eine Kopie der Daten statt für DynamoDB ausführen. Dies spart Lesevorgänge und schont den bereitgestellten Durchsatz.

Kinesis

EMRAmazon-Cluster können Amazon Kinesis-Streams direkt lesen und verarbeiten, indem sie vertraute Tools aus dem Hadoop-Ökosystem wie Hive, Pig MapReduce, Hadoop Streaming API und Cascading verwenden. Sie können auch Echtzeitdaten von Amazon Kinesis mit vorhandenen Daten auf Amazon S3, Amazon DynamoDB und HDFS in einem laufenden Cluster verbinden. Sie können die Daten für die Nachbearbeitung direkt von Amazon EMR in Amazon S3 oder DynamoDB laden. Informationen zu den wichtigsten Services und Preisen von Amazon Kinesis finden Sie auf der Seite zu [Amazon Kinesis](#).

Was kann ich mit der Integration von Amazon EMR und Amazon Kinesis machen?

Die Integration zwischen Amazon EMR und Amazon Kinesis macht bestimmte Szenarien viel einfacher, zum Beispiel:

- Streaming-Protokollanalyse – Sie können Streaming-Webprotokolle analysieren, um eine Liste der 10 häufigsten Fehlertypen alle „x“ Minuten nach Region, Browser und Zugriffs-Domain zu erzeugen.
- Kundenbindung – Sie können Abfragen schreiben, die Clickstream-Daten aus Amazon Kinesis mit Informationen zu Werbekampagnen in einer DynamoDB-Tabelle verknüpfen, um die effektivsten Kategorien von Anzeigen zu bestimmen, die auf bestimmten Websites angezeigt werden.
- Interaktive Ad-hoc-Abfragen — Sie können regelmäßig Daten aus Amazon Kinesis Kinesis-Streams in eine lokale Impala-Tabelle laden HDFS und diese für schnelle, interaktive Analyseabfragen verfügbar machen.

Analyse von Amazon-Kinesis-Streams mit Checkpointing

Benutzer können regelmäßige, auf Stapelverarbeitung basierende Analysen von Amazon-Kinesis-Streams in so genannten Iterationen ausführen. Da Amazon Kinesis Kinesis-Stream-Datensätze

mithilfe einer Sequenznummer abgerufen werden, werden Iterationsgrenzen durch Start- und Endsequenznummern definiert, die Amazon in einer DynamoDB-Tabelle EMR speichert. Wenn z. B. `iteration0` endet, wird die letzte Sequenznummer in DynamoDB gespeichert, sodass bei Beginn des Auftrags `iteration1` nachfolgende Daten aus dem Stream abgerufen werden können. Diese Zuweisung von Iterationen in Stream-Daten wird als Checkpointing bezeichnet. Weitere Informationen finden Sie unter [Kinesis-Konnektor](#).

Wenn eine Iteration mit einem Checkpoint versehen wurde und der Job eine Iteration nicht verarbeiten konnte, EMR versucht Amazon, die Datensätze in dieser Iteration erneut zu verarbeiten.

Checkpointing ist eine Funktion, mit der Sie Folgendes durchführen können:

- Starten von Datenverarbeitung nach einer Sequenznummer, die von einer vorherigen Abfrage verarbeitet wurde, die im selben Stream und mit selben logischen Namen ausgeführt wurde
- Erneutes Verarbeiten desselben Datenstapels aus Kinesis, der von einer früheren Abfrage verarbeitet wurde

Setzen Sie in Ihren Skripten den Parameter `kinesis.checkpoint.enabled` auf `true`, um Checkpointing zu aktivieren. Konfigurieren Sie außerdem die folgenden Parameter:

| Konfigurationseinstellung | Beschreibung |
|---|---|
| <code>kinesis.checkpoint.metastore.table.name</code> | Name der DynamoDB-Tabelle, in der Checkpoint-Informationen gespeichert werden |
| <code>kinesis.checkpoint.metastore.hash.key.name</code> | Name des Hash-Schlüssels für die DynamoDB-Tabelle |
| <code>kinesis.checkpoint.metastore.hash.range.name</code> | Name des Bereichsschlüssels für die DynamoDB-Tabelle |
| <code>kinesis.checkpoint.logical.name</code> | Ein logischer Name für die aktuelle Verarbeitung |
| <code>kinesis.checkpoint.iteration.no</code> | Iterationsnummer für die Verarbeitung im Zusammenhang mit dem logischen Namen |
| <code>kinesis.rerun.iteration.without.wait</code> | Boolescher Wert, der angibt, ob eine fehlgeschlagene Iteration erneut ausgeführt werden kann, |

| Konfigurationseinstellung | Beschreibung |
|---------------------------|---|
| | ohne die Zeitbeschränkung abzuwarten. Der Standardwert ist <code>false</code> |

Bereitgestellte IOPS Empfehlungen für Amazon DynamoDB-Tabellen

Der EMR Amazon-Connector für Amazon Kinesis verwendet die DynamoDB-Datenbank als Grundlage für Checkpointing-Metadaten. Sie müssen eine Tabelle in DynamoDB erstellen, bevor Sie Daten in einem Amazon Kinesis Kinesis-Stream mit einem EMR Amazon-Cluster in Checkpoint-Intervallen verwenden können. Die Tabelle muss sich in derselben Region wie Ihr EMR Amazon-Cluster befinden. Im Folgenden finden Sie allgemeine Empfehlungen für die Anzahl von, die IOPS Sie für Ihre DynamoDB-Tabellen bereitstellen sollten. Geben Sie die maximale Anzahl von Hadoop-Jobs (mit unterschiedlicher Kombination aus logischem Namen und Iterationsnummer) an, die gleichzeitig ausgeführt werden können, und die maximale Anzahl von Shards, die ein Job verarbeiten kann: `j`

Für Read Capacity Units (Lese-Kapazitätseinheiten): `j*s/5`

Für Write Capacity Units (Schreibkapazitätseinheiten): `j*s`

Leistungsaspekte

Der Amazon Kinesis Kinesis-Shard-Durchsatz ist direkt proportional zur Instance-Größe der Knoten in EMR Amazon-Clustern und der Datensatzgröße im Stream. Wir empfehlen die Verwendung von `m5.xlarge` oder größeren Instances für Haupt- und Core-Knoten.

Planen Sie die Amazon Kinesis Kinesis-Analyse mit Amazon EMR

Wenn Sie Daten in einem aktiven Amazon-Kinesis-Stream analysieren und dabei durch Timeouts und eine maximale Dauer der einzelnen Iterationen eingeschränkt sind, müssen Sie die Analyse unbedingt häufiger durchführen, um regelmäßig Details aus dem Stream zu erhalten. Es gibt verschiedene Möglichkeiten, solche Skripte und Abfragen in regelmäßigen Intervallen durchzuführen. Wir empfehlen für wiederkehrende Aufgaben wie diese [AWS Data Pipeline](#). Weitere Informationen finden Sie unter [AWS Data Pipeline PigActivity](#) und [AWS Data Pipeline HiveActivity](#) im AWS Data Pipeline Entwicklerhandbuch.

Migration des Spark Kinesis-Connectors auf SDK 2.x für Amazon 7.0 EMR

Das AWS SDK bietet eine Vielzahl von Bibliotheken für die Interaktion mit AWS Cloud-Computing-Diensten, z. B. die Verwaltung von Anmeldeinformationen APIs und die Verbindung zu S3- und Kinesis-Diensten. Der Spark-Kinesis-Konnektor wird verwendet, um Daten aus Kinesis Data Streams zu verarbeiten, und die empfangenen Daten werden in der Ausführungs-Engine von Spark transformiert und verarbeitet. Derzeit baut dieser Konnektor auf 1.x von AWS SDK und Kinesis-client-library (KCL) auf.

Im Rahmen der AWS SDK 2.x-Migration wird auch der Spark Kinesis-Connector entsprechend aktualisiert, sodass er mit 2.x ausgeführt werden kann. In der Amazon EMR 7.0-Version enthält Spark das SDK 2.x-Upgrade, das in der Community-Version von Apache Spark noch nicht verfügbar ist. Wenn Sie den Spark Kinesis-Connector aus einer Version unter 7.0 verwenden, müssen Sie Ihre Anwendungscodes für die Ausführung auf SDK 2.x migrieren, bevor Sie zu Amazon EMR 7.0 migrieren können.

Migrationshandbücher

In diesem Abschnitt werden die Schritte zur Migration einer Anwendung zum aktualisierten Spark-Kinesis-Konnektor beschrieben. Es enthält Anleitungen für die Migration zur Kinesis Client Library (KCL) 2.x, AWS Anmeldeinformationsanbieter und AWS Service-Clients in 2.x. Als Referenz enthält es auch ein [WordCount](#) Beispielprogramm, das den Kinesis-Konnektor verwendet.

Themen

- [Migration KCL von 1.x zu 2.x](#)
- [Migrieren von Anbietern von Anmeldeinformationen von 1.x auf 2.x AWS SDK](#)
- [Migration von AWS Service-Clients von AWS SDK 1.x auf 2.x](#)
- [Codebeispiele für Streaming-Anwendungen](#)
- [Überlegungen zur Verwendung des aktualisierten Spark-Kinesis-Konnektors](#)

Migration KCL von 1.x zu 2.x

- Ebene und Dimensionen der Metriken in **KinesisInputDStream**

Wenn Sie einen `KinesisInputDStream` instanziiieren, können Sie die Metrikebene und die Dimensionen für den Stream steuern. Das folgende Beispiel zeigt, wie Sie diese Parameter mit 1.x anpassen können: KCL

```
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()
```

In KCL 2.x haben diese Konfigurationseinstellungen unterschiedliche Paketnamen. Für die Migration zu 2.x:

1. Ändern Sie die Importanweisungen für `com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration` und `com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel` zu `software.amazon.kinesis.metrics.MetricsLevel` bzw. `software.amazon.kinesis.metrics.MetricsUtil`.

```
// 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. Ersetzen Sie die Zeile `metricsEnabledDimensionsKinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS` durch `metricsEnabledDimensionsSet(MetricsUtil.OPERATION_DIMENSION_NAME, MetricsUtil.SHARD_ID_DIMENSION_NAME)`

Im Folgenden finden Sie eine aktualisierte Version von `KinesisInputDStream` mit benutzerdefinierten Metrikebene und Metrikdimensionen:

```
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()
```

- Meldungshandler-Funktion in `KinesisInputDStream`

Bei der Instanziierung eines `KinesisInputDStream` können Sie auch eine „Meldungshandler-Funktion“ angeben, die einen Kinesis-Datensatz verwendet und ein generisches Objekt `T` zurückgibt, falls Sie andere in einem Datensatz enthaltene Daten wie den Partitionsschlüssel verwenden möchten.

In KCL 1.x lautet die Signatur der Nachrichtenhandler-Funktion: `Record => T`, wobei `Record` steht. `com.amazonaws.services.kinesis.model.Record` In KCL 2.x wurde die Signatur des Handlers in: `KinesisClientRecord => T`, wo `KinesisClientRecord` ist, geändert.
`software.amazon.kinesis.retrieval.KinesisClientRecord`

Es folgt ein Beispiel für die Bereitstellung eines Message-Handlers in KCL 1.x:

```
import com.amazonaws.services.kinesis.model.Record

def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
```



```

.streamName(streamName)
.endpointUrl(endpointUrl)
.regionName(regionName)
.initialPosition(new Latest())
.checkpointAppName(appName)
.checkpointInterval(Seconds(10))
.storageLevel(StorageLevel.MEMORY_ONLY)
.buildWithMessageHandler(addFive)

```

Für die Migration des Meldungshandlers:

1. Ändern Sie die Importanweisung für `com.amazonaws.services.kinesis.model.Record` zu `software.amazon.kinesis.retrieval.KinesisClientRecord`.

```

// import com.amazonaws.services.kinesis.model.Record
import software.amazon.kinesis.retrieval.KinesisClientRecord

```

2. Aktualisieren Sie die Methodensignatur des Meldungshandlers.

```

//def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt
+ 5

```

Es folgt ein aktualisiertes Beispiel für die Bereitstellung des Message-Handlers in KCL 2.x:

```

import software.amazon.kinesis.retrieval.KinesisClientRecord

def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
  .storageLevel(StorageLevel.MEMORY_ONLY)
  .buildWithMessageHandler(addFive)

```

Weitere Informationen zur Migration von KCL 1.x auf 2.x finden Sie unter [Migration von Verbrauchern](#) von 1.x auf 2.x. KCL KCL

Migrieren von Anbietern von Anmeldeinformationen von 1.x auf 2.x AWS SDK

Anbieter von Anmeldeinformationen werden verwendet, um AWS Anmeldeinformationen für Interaktionen mit zu erhalten. AWS In SDK 2.x gibt es mehrere Schnittstellen- und Klassenänderungen im Zusammenhang mit den Anbietern von Anmeldeinformationen, die Sie [hier finden](#). Der Spark Kinesis-Konnektor hat eine Schnittstelle (`org.apache.spark.streaming.kinesis.SparkAWSCredentials`) und Implementierungsklassen definiert, die die Version 1.x von AWS Credential Providern zurückgeben. Diese Anbieter von Anmeldeinformationen werden bei der Initialisierung von Kinesis-Clients benötigt. Wenn Sie die Methode beispielsweise `SparkAWSCredentials.provider` in den Anwendungen verwenden, müssten Sie die Codes aktualisieren, um die 2.x-Version der Credential Provider nutzen zu können. AWS

Im Folgenden finden Sie ein Beispiel für die Verwendung der Anmeldeinformationsanbieter in 1.x: AWS SDK

```
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")
```

Um zu 2.x zu migrieren SDK:

1. Ändern Sie die Importanweisung für `com.amazonaws.auth.AWSCredentialsProvider` zu `software.amazon.awssdk.auth.credentials.AwsCredentialsProvider`

```
//import com.amazonaws.auth.AWSCredentialsProvider
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

2. Aktualisieren Sie die verbleibenden Codes, die diese Klasse verwenden.

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert (credentialProvider.isInstanceOf[AwsCredentialsProvider], "Type should be
    AwsCredentialsProvider")
```

Migration von AWS Service-Clients von AWS SDK 1.x auf 2.x

AWS Service-Clients haben in 2.x unterschiedliche Paketnamen (d. h. `software.amazon.awssdk`), wohingegen in Version 1.x andere Paketnamen verwendet werden. SDK.com.amazonaws Weitere Informationen über die Änderungen in dieser Version finden Sie [hier](#). Wenn Sie diese Service-Clients in den Codes verwenden, müssten Sie die Clients entsprechend migrieren.

Es folgt ein Beispiel für die Erstellung eines Clients in SDK 1.x:

```
import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
import com.amazonaws.services.dynamodbv2.document.DynamoDB

AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

Für die Migration zu 2.x:

1. Ändern Sie die Importanweisungen für Service-Clients. Nehmen wir als Beispiel DynamoDB-Clients. Sie müssten `com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient` oder `com.amazonaws.services.dynamodbv2.document.DynamoDB` zu `software.amazon.awssdk.services.dynamodb.DynamoDbClient` ändern.

```
// import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
// import com.amazonaws.services.dynamodbv2.document.DynamoDB
import software.amazon.awssdk.services.dynamodb.DynamoDbClient
```

2. Die Codes aktualisieren, die die Clients initialisieren

```
// AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
// AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();

DynamoDbClient ddbClient = DynamoDbClient.create();
DynamoDbClient ddbClient = DynamoDbClient.builder().build();
```

Weitere Informationen zur Migration AWS SDK von 1.x zu 2.x finden Sie unter [Was ist der Unterschied zwischen dem AWS SDK für Java 1.x und 2.x](#)

Codebeispiele für Streaming-Anwendungen

```
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()
}
}

```

Überlegungen zur Verwendung des aktualisierten Spark-Kinesis-Konnektors

- Wenn Ihre Anwendungen die JDK Version `kinesis-producer-library` mit einer niedrigeren Version als 11 verwenden, können Sie auf Ausnahmen stoßen wie. `java.lang.NoClassDefFoundError: javax/xml/bind/DatatypeConverter` Dies liegt daran, dass EMR 7.0 standardmäßig JDK 17 enthält und J2EE-Module seit Java 11+ aus den Standardbibliotheken entfernt wurden. Dies könnte behoben werden, indem die folgende Abhängigkeit zur POM-Datei hinzugefügt wird. Ersetzen Sie die Bibliotheksversion nach Bedarf durch eine passende.

```
<dependency>
  <groupId>javax.xml.bind</groupId>
  <artifactId>jaxb-api</artifactId>
  <version>${jaxb-api.version}</version>
</dependency>
```

- Das Spark Kinesis-Connector-JAR befindet sich nach der Erstellung eines EMR Clusters unter diesem Pfad: `/usr/lib/spark/connector/lib/`

S3 DistCp (s3-dist-cp)

Apache DistCp ist ein Open-Source-Tool, mit dem Sie große Datenmengen kopieren können. S3 DistCp ähnelt Amazon S3 DistCp, ist aber für die Verwendung mit AWS Amazon S3 optimiert. Der Befehl für S3 DistCp in Amazon EMR Version 4.0 und höher lautet `s3-dist-cp`, den Sie als Schritt in einem Cluster oder in der Befehlszeile hinzufügen. Mit S3 DistCp können Sie große Datenmengen effizient von Amazon S3 dorthin kopieren, HDFS wo sie in nachfolgenden Schritten in Ihrem EMR Amazon-Cluster verarbeitet werden können. Sie können S3 auch verwenden `DistCp`, um Daten zwischen Amazon S3-Buckets oder von HDFS zu Amazon S3 zu kopieren. S3 DistCp ist skalierbarer und effizienter für das parallel Kopieren einer großen Anzahl von Objekten über Buckets und AWS Konten hinweg.

Spezifische Befehle, die die Flexibilität von S3DistCP in realen Szenarien demonstrieren, finden Sie im Big Data-Blog unter [Sieben Tipps zur Verwendung von S3 DistCp](#). AWS

Wie bei DistCp S3 DistCp wird häufig MapReduce auf verteilte Weise kopiert. Kopier-, Fehlerbehandlungs-, Wiederherstellungs- und Berichterstellungsaufgaben werden auf mehrere Server verteilt. Weitere Informationen zum DistCp Open-Source-Projekt Apache finden Sie in der [DistCpAnleitung](#) in der Apache Hadoop-Dokumentation.

Wenn S3 DistCp einige oder alle der angegebenen Dateien nicht kopieren kann, schlägt der Cluster-Schritt fehl und gibt einen Fehlercode ungleich Null zurück. In diesem Fall bereinigt S3 teilweise kopierte Dateien DistCp nicht.

Important

S3 unterstützt DistCp keine Amazon S3 S3-Bucket-Namen, die den Unterstrich enthalten.

S3 unterstützt DistCp keine Verkettung von Parquet-Dateien. Verwenden Sie stattdessen PySpark. Weitere Informationen finden Sie unter [Koncatenating Parquet-Dateien in Amazon EMR](#).


Verwenden Sie EMR Amazon-Version 5.33.0 oder höher oder Amazon-Version 6.3.0 oder höher, um Kopierfehler zu vermeiden, wenn Sie S3DistCP verwenden. Verwenden Sie HDFS, um eine einzelne Datei (anstelle eines Verzeichnisses) von S3 nach S3 zu kopieren.

DistCp S3-Optionen

S3 ist zwar ähnlich DistCp, DistCp unterstützt aber verschiedene Optionen, mit denen Sie die Art und Weise ändern können, wie Daten kopiert und komprimiert werden.

Wenn Sie S3 aufrufen DistCp, können Sie die in der folgenden Tabelle beschriebenen Optionen angeben. Die Optionen werden dem Schritt mithilfe der Liste der Argumente hinzugefügt. Beispiele für die DistCp S3-Argumente sind in der folgenden Tabelle aufgeführt.

Option	Beschreibung	Erforderlich
<code>--src=LOCATION</code>	<p>Der Speicherort der zu kopierenden Daten. Dies kann entweder ein HDFS oder ein Amazon S3 S3-Standort sein.</p> <p>Beispiel: <code>--src=s3://amzn-s3-demo-bucket1/logs/j-3GYXXXXXX9I0J/node</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>⚠ Important</p> <p>S3 unterstützt DistCp keine Amazon S3 S3-Bucket-Namen, die den Unterstrich enthalten.</p> </div>	Ja
<code>--dest=LOCATION</code>	<p>Das Ziel für die Daten. Dies kann entweder ein HDFS oder ein Amazon S3 S3-Standort sein.</p> <p>Beispiel: <code>--dest=hdfs:///output</code></p>	Ja

Option	Beschreibung	Erforderlich
	<div style="border: 1px solid #f08080; padding: 10px; background-color: #fff9e6;"> <p> Important</p> <p>S3 unterstützt DistCp keine Amazon S3 S3-Bucket-Namen, die den Unterstrich enthalten.</p> </div>	
<pre>--srcPattern=PATTERN</pre>	<p>Ein regulärer Ausdruck, der die Kopieroperation auf eine Teilmenge der Daten unter <code>--src</code> filtert. Wenn weder <code>--srcPattern</code> noch <code>--groupBy</code> angegeben ist, werden alle Daten unter <code>--src</code> in <code>--dest</code> kopiert.</p> <p>Wenn das Argument des regulären Ausdrucks Sonderzeichen wie Sternchen (*) enthält, muss entweder der reguläre Ausdruck oder die gesamte <code>--args</code>-Zeichenfolge in einfache Anführungszeichen (') gesetzt werden.</p> <p>Beispiel: <code>--srcPattern=.*daemons.*-hadoop-.*</code></p>	Nein

Option	Beschreibung	Erforderlich
<pre>--groupBy=PATTERN</pre>	<p>Ein regulärer Ausdruck, der S3 veranlasst, Dateien DistCp zu verketteten, die dem Ausdruck entsprechen. Sie können mit dieser Option beispielsweise alle Protokolldateien kombinieren, die innerhalb einer Stunde in einer einzigen Datei geschrieben wurden. Der verkettete Dateiname ist der Wert, der dem regulären Ausdruck für die Gruppe entspricht.</p> <p>Klammern geben an, wie Dateien gruppiert werden soll, wobei alle Elemente, die der eingeklammerten Anweisung entsprechen, zu einer einzigen Ausgabedatei kombiniert werden. Wenn der reguläre Ausdruck keine Anweisung in Klammern enthält, schlägt der Cluster beim Schritt S3 DistCp fehl und gibt einen Fehler zurück.</p> <p>Wenn das Argument des regulären Ausdrucks Sonderzeichen wie Sternchen (*) enthält, muss entweder der reguläre Ausdruck oder die gesamte <code>--args</code>-Zeichenfolge in einfache Anführungszeichen (') gesetzt werden.</p> <p>Wenn <code>--groupBy</code> angegeben ist, werden nur Dateien kopiert, die dem angegebenen Muster entsprechen. Sie müssen <code>--groupBy</code> und <code>--srcPattern</code> nicht gleichzeitig angeben.</p> <p>Beispiel: <code>--groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+-[0-9]+)</code>.</p>	Nein

Option	Beschreibung	Erforderlich
<code>--targetSize=SIZE</code>	<p>Die Größe (in Mebibyte (MiB)) der Dateien, die basierend auf der <code>--groupBy</code> -Option erstellt werden sollen. Dieser Wert muss eine Ganzzahl sein. Wenn dieser <code>--targetSize</code> Wert gesetzt ist, DistCp versucht S3, dieser Größe zu entsprechen. Die tatsächliche Größe der kopierten Dateien kann größer oder kleiner als dieser Wert sein. Aufträge werden basierend auf der Größe der Datendatei aggregiert. Deshalb ist es möglich, dass die Zieldateigröße der Quelldateigröße entspricht.</p> <p>Wenn die durch <code>--groupBy</code> verketteten Dateien größer als der Wert <code>--targetSize</code> sind, werden sie in Teildateien aufgelöst und sequenziell mit einem an das Ende angefügten numerischen Wert benannt. Eine Datei, die beispielsweise zu <code>myfile.gz</code> verkettet ist, wird wie folgt in mehrere Teile unterteilt: <code>myfile0.gz</code> , <code>myfile1.gz</code> usw.</p> <p>Beispiel: <code>--targetSize=2</code></p>	Nein
<code>--appendToLastFile</code>	<p>Gibt das Verhalten von S3 DistCp beim Kopieren in Dateien von Amazon S3 anHDFS, in die bereits vorhanden sind. Es werden neue Dateidaten an vorhandene Dateien angefügt. Wenn Sie <code>--appendToLastFile</code> mit <code>--groupBy</code> verwenden, werden neue Daten an Dateien angefügt, die denselben Gruppen entsprechen. Diese Option respektiert auch das <code>--targetSize</code> -Verhalten, wenn sie mit <code>--groupBy</code> verwendet wird.</p>	Nein

Option	Beschreibung	Erforderlich
<code>--outputCodec=CODEC</code>	<p>Gibt den Komprimierungs-Codec für die kopierten Dateien an. Mögliche Werte sind: <code>gzip</code>, <code>gz</code>, <code>lzo</code>, <code>snappy</code> oder <code>none</code>. Sie können diese Option verwenden, um beispielsweise mit Gzip komprimierte Eingabedateien in komprimierte Ausgabedateien zu konvertieren oder um die Dateien als Teil des Kopiervorgangs zu dekomprimieren. LZO</p> <p>Wenn Sie einen Ausgabe-Codec auswählen, wird dem Dateinamen die entsprechende Erweiterung angefügt (z. B. für <code>gz</code> und <code>gzip</code> die Erweiterung <code>.gz</code>). Wenn Sie keinen Wert für <code>--outputCodec</code> angeben, werden die Dateien ohne Änderung der Komprimierung kopiert.</p> <p>Beispiel: <code>--outputCodec=lzo</code></p>	Nein
<code>--s3ServerSideEncryption</code>	<p>Stellt sicher, dass die Zieldaten mithilfe eines AWS serviceseitigen Schlüssels übertragen werden und in Amazon S3 automatisch verschlüsselt werden. Beim Abrufen von Daten mit S3 DistCp werden die Objekte automatisch entschlüsselt. Wenn Sie versuchen, ein entschlüsseltes Objekt in einen Amazon-S3-Bucket zu kopieren, für den eine Verschlüsselung erforderlich ist, schlägt die Operation fehl. Weitere Informationen finden Sie unter Verwenden der Datenverschlüsselung.</p> <p>Beispiel: <code>--s3ServerSideEncryption</code></p>	Nein

Option	Beschreibung	Erforderlich
<code>--deleteOnSuccess</code>	<p>Wenn der Kopiervorgang erfolgreich ist, veranlasst diese Option S3, die kopierten Dateien vom Quellspeicherort DistCp zu löschen. Dies ist nützlich, wenn Sie Ausgabedateien, wie Protokolldateien, von einem Speicherort an einen anderen als geplante Aufgabe kopieren und dieselben Dateien nicht zweimal kopiert werden sollen.</p> <p>Beispiel: <code>--deleteOnSuccess</code></p>	Nein
<code>--disableMultipartUpload</code>	<p>Deaktiviert die Verwendung des mehrteiligen Uploads.</p> <p>Beispiel: <code>--disableMultipartUpload</code></p>	Nein
<code>--multipartUploadChunkSize=SIZE</code>	<p>Die Größe jedes Teils in einem mehrteiligen Amazon-S3-Upload in MiB. S3 DistCp verwendet den mehrteiligen Upload, wenn Daten kopiert werden, die <code>multipartUploadChunkSize</code> größer sind als. Um die Arbeitsleistung zu verbessern, können Sie die Größe der einzelnen Teile erhöhen. Die Standardgröße beträgt 128 MiB.</p> <p>Beispiel: <code>--multipartUploadChunkSize=1000</code></p>	Nein
<code>--numberFiles</code>	<p>Stellt Ausgabedateien sequenzielle Nummern voran. Die Zählung beginnt bei 0, es sei denn, von <code>--startingIndex</code> wird ein anderer Wert angegeben.</p> <p>Beispiel: <code>--numberFiles</code></p>	Nein

Option	Beschreibung	Erforderlich
<code>--startingIndex=INDEX</code>	<p>Wird mit <code>--numberOfFiles</code> verwendet, um die erste Nummer in der Sequenz anzugeben.</p> <p>Beispiel: <code>--startingIndex=1</code></p>	Nein
<code>--outputManifest=FILENAME</code>	<p>Erstellt eine mit Gzip komprimierte Textdatei, die eine Liste aller von S3 kopierten Dateien enthält.</p> <p>DistCp</p> <p>Beispiel: <code>--outputManifest=manifest-1.gz</code></p>	Nein
<code>--previousManifest=PATH</code>	<p>Liest eine Manifestdatei, die bei einem früheren Aufruf von S3 DistCp mit dem <code>--outputManifest</code> Flag erstellt wurde. Wenn das <code>--previousManifest</code> Flag gesetzt ist, DistCp schließt S3 die im Manifest aufgelisteten Dateien vom Kopiervorgang aus. Wenn <code>--outputManifest</code> zusammen mit <code>--previousManifest</code> angegeben ist, erscheinen die im vorherigen Manifest aufgeführten Dateien auch in der neuen Manifestdatei, obwohl die Dateien nicht kopiert werden.</p> <p>Beispiel: <code>--previousManifest=/usr/bin/manifest-1.gz</code></p>	Nein
<code>--requirePreviousManifest</code>	<p>Erfordert ein vorheriges Manifest, das während eines vorherigen Aufrufs von S3 DistCp erstellt wurde. Wenn dieser Wert auf "false" festgelegt ist, wird kein Fehler generiert, wenn kein vorheriges Manifest angegeben ist. Der Standardwert ist „true“.</p>	Nein

Option	Beschreibung	Erforderlich
<code>--copyFromManifest</code>	<p>Kehrt das Verhalten <code>--previousManifest</code> um, dass S3 DistCp die angegebene Manifestdatei als Liste der zu kopierenden Dateien verwendet, anstatt als Liste von Dateien, die vom Kopieren ausgeschlossen werden sollen.</p> <p>Beispiel: <code>--copyFromManifest --previousManifest=/usr/bin/manifest-1.gz</code></p>	Nein
<code>--s3Endpoint=ENDPOINT</code>	<p>Gibt den Amazon-S3-Endpoint an, der beim Hochladen einer Datei verwendet werden soll. Mit dieser Option wird der Endpoint sowohl für die Quelle als auch das Ziel festgelegt. Falls die Option nicht festgelegt ist, wird als Standardendpoint <code>s3.amazonaws.com</code> verwendet. Eine Liste der Amazon-S3-Endpoints finden Sie unter Regionen und Endpunkte.</p> <p>Beispiel: <code>--s3Endpoint=s3.eu-west-1.amazonaws.com</code></p>	Nein
<code>--storageClass=CLASS</code>	<p>Die zu verwendende Speicherklasse, wenn das Ziel Amazon S3 ist. Gültige Werte sind <code>STANDARD</code> und <code>REDUCED_REDUNDANCY</code>. Wenn diese Option nicht angegeben ist, DistCp versucht S3, die Speicherklasse beizubehalten.</p> <p>Beispiel: <code>--storageClass=STANDARD</code></p>	Nein

Option	Beschreibung	Erforderlich
<code>--srcPrefixesFile=PATH</code>	<p>eine Textdatei in Amazon S3 (s3://), HDFS (hdfs://) oder einem lokalen Dateisystem (file:/), die eine Liste von <code>src</code> Präfixen enthält, ein Präfix pro Zeile.</p> <p>Wenn <code>srcPrefixesFile</code> angegeben, listet S3 den <code>DistCp</code> <code>src</code>-Pfad nicht auf. Stattdessen wird eine Quellliste als kombiniertes Ergebnis der Auflistung aller in dieser Datei angegebenen Präfixe erstellt. Der relative Pfad im Vergleich zum <code>src</code>-Pfad, wird statt dieser Präfixe verwendet, um die Zielpfade zu generieren. Wenn <code>srcPattern</code> angegeben ist, wird dieser Wert auf die kombinierten Listenergebnisse der Quellpräfixe angewendet, um die Eingabe weiter zu filtern. Wenn <code>copyFromManifest</code> verwendet wird, werden die Objekte im Manifest kopiert und <code>srcPrefixesFile</code> wird ignoriert.</p> <p>Beispiel: <code>--srcPrefixesFile=PATH</code></p>	Nein

Zusätzlich zu den oben genannten Optionen `DistCp` implementiert S3 die [Tool-Schnittstelle](#), was bedeutet, dass es die generischen Optionen unterstützt.

S3 `DistCp` als Schritt in einem Cluster hinzufügen

Sie können S3 aufrufen, `DistCp` indem Sie es als Schritt zu Ihrem Cluster hinzufügen. Schritte können einem Cluster beim Start oder einem laufenden Cluster mithilfe der Konsole CLI, oder hinzugefügt API werden. Die folgenden Beispiele veranschaulichen das Hinzufügen eines `DistCp` S3-Schritts zu einem laufenden Cluster. Weitere Informationen zum Hinzufügen von Schritten zu einem Cluster finden Sie unter [Arbeit an einem Cluster einreichen](#) im Amazon EMR Management Guide.

Um einem laufenden Cluster einen `DistCp` S3-Schritt hinzuzufügen, verwenden Sie den AWS CLI

Weitere Informationen zur Verwendung von EMR Amazon-Befehlen in der AWS CLI finden Sie in der [AWS CLI Befehlsreferenz](#).

- Um einem Cluster, der S3 aufruft, einen Schritt hinzuzufügen `DistCp`, übergeben Sie die Parameter, die angeben, wie S3 den Kopiervorgang ausführen `DistCp` soll, als Argumente.

Im folgenden Beispiel werden Daemon-Protokolle von Amazon S3 in `hdfs:///output` kopiert. Im folgenden Befehl:

- `--cluster-id` gibt den Cluster an.
- `Jar` ist der Speicherort der `DistCp` JAR S3-Datei. Ein Beispiel für die Ausführung eines Befehls auf einem Cluster mithilfe von `command-runner.jar` finden Sie unter [Senden eines benutzerdefinierten JAR Schritts zur Ausführung eines Skripts oder Befehls](#).
- `Args` ist eine durch Kommas getrennte Liste der Optionsname-Wert-Paare, die an S3 übergeben werden sollen. `DistCp` Eine vollständige Liste der verfügbaren Optionen finden Sie unter [DistCp S3-Optionen](#).

Um einem laufenden Cluster einen `DistCp` S3-Kopierschritt hinzuzufügen, fügen Sie wie *myStep.json* in diesem Beispiel Folgendes in eine JSON Datei ein, die in Amazon S3 oder Ihrem lokalen Dateisystem gespeichert ist. Ersetzen `j-3GYXXXXXX9I0K` mit Ihrer Cluster-ID und ersetzen *mybucket* mit Ihrem Amazon S3 S3-Bucket-Namen.

```
[
  {
    "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 Kopieren Sie Protokolldateien von Amazon S3 nach HDFS

Dieses Beispiel zeigt auch, wie Protokolldateien, die in einem Amazon S3 S3-Bucket gespeichert sind, kopiert werden, HDFS indem ein Schritt zu einem laufenden Cluster hinzugefügt wird. In

diesem Beispiel werden die Daten, die in die Daemon-Protokolle kopiert werden, anhand der Option `--srcPattern` eingeschränkt.

Um Protokolldateien von Amazon S3 zu kopieren, um die `--srcPattern` Option zu HDFS verwenden, fügen Sie Folgendes in eine JSON Datei ein, die in Amazon S3 oder Ihrem lokalen Dateisystem gespeichert ist, wie in diesem *myStep.json* Beispiel gezeigt. Ersetzen *j-3GYXXXXXX9I0K* mit Ihrer Cluster-ID und ersetzen Sie *mybucket* mit Ihrem Amazon S3 S3-Bucket-Namen.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*daemons.*-hadoop-.*"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

Aufräumen nach fehlgeschlagenen DistCp S3-Jobs

Wenn S3 einige oder alle der angegebenen Dateien DistCp nicht kopieren kann, schlägt der Befehl oder der Clusterschritt fehl und gibt einen Fehlercode ungleich Null zurück. In diesem Fall bereinigt S3 teilweise kopierte Dateien DistCp nicht. Sie müssen sie manuell löschen.

Teilweise kopierte Dateien werden in dem HDFS tmp Verzeichnis in Unterverzeichnissen mit der eindeutigen Kennung des DistCp S3-Jobs gespeichert. Die ID finden Sie in der Standardausgabe der Aufgabe.

Für einen DistCp S3-Job mit der ID können Sie `4b1c37bb-91af-4391-aaaf8-46a6067085a6` beispielsweise eine Verbindung zum Master-Knoten des Clusters herstellen und den folgenden Befehl ausführen, um die mit dem Job verknüpften Ausgabedateien anzuzeigen.

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aaaf8-46a6067085a6/output
```

Der Befehl gibt eine Liste von Dateien ähnlich der folgenden zurück:

```
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
```

Sie können dann den folgenden Befehl ausführen, um das Verzeichnis und den gesamten Inhalt zu löschen.

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```

Befehle und Skripte auf einem Amazon-EMR-Cluster ausführen

In diesem Thema wird beschrieben, wie Sie einen Befehl oder ein Skript als Schritt auf Ihrem Cluster ausführen. Das Ausführen eines Befehls oder Skripts als Schritt ist eine der vielen Möglichkeiten, wie Sie [Arbeit an einen Cluster weiterleiten](#) können. Dies ist in den folgenden Situationen hilfreich:

- Wenn Sie keinen SSH-Zugriff auf Ihren Amazon-EMR-Cluster haben
- Wenn Sie einen Bash- oder Shell-Befehl zur Fehlerbehebung in Ihrem Cluster ausführen möchten

Sie können ein Skript ausführen, wenn Sie einen Cluster erstellen oder wenn sich Ihr Cluster im WAITING-Status befindet. Um ein Skript auszuführen, bevor die Schrittverarbeitung beginnt, verwenden Sie stattdessen eine Bootstrap-Aktion. Weitere Informationen finden Sie unter [Bootstrap-Aktionen zum Installieren zusätzlicher Software erstellen](#) im Verwaltungshandbuch für Amazon EMR.

Amazon EMR bietet die folgenden Tools, mit denen Sie Skripts, Befehle und andere Cluster-Programme ausführen können. Sie können beide Tools über die Amazon-EMR-Managementkonsole oder die AWS CLI aufrufen.

`command-runner.jar`

Befindet sich auf dem Amazon-EMR-AMI für Ihren Cluster. Sie können `command-runner.jar` verwenden, um Befehle auf Ihrem Cluster auszuführen. Sie geben den Pfad `command-runner.jar` an, ohne den vollständigen Pfad zu verwenden.

`script-runner.jar`

Gehostet auf Amazon S3 auf `s3://<region>.elasticmapreduce/libs/script-runner/script-runner.jar`, wobei `<region>` die Region ist, in der sich Ihr Amazon-EMR-Cluster befindet. Sie können `script-runner.jar` nutzen, um damit lokal oder auf Amazon S3 gespeicherte Skripte auf Ihrem Cluster auszuführen. Sie müssen die vollständige URI auf `script-runner.jar` angeben, wenn Sie einen Schritt einreichen.

Reichen Sie einen benutzerdefinierten JAR-Schritt ein, um ein Skript oder einen Befehl auszuführen

Die folgenden AWS CLI Beispiele veranschaulichen einige gängige Anwendungsfälle von `command-runner.jar` und `script-runner.jar` auf Amazon EMR.

Example : Ausführen eines Befehls in einem Cluster mit **command-runner.jar**

Wenn Sie `command-runner.jar` verwenden, geben Sie Befehle, Optionen und Werte in der Argumentliste Ihres Schritts an.

Im folgenden AWS CLI-Beispiel wird ein Schritt an einen aktiven Cluster gesendet, der `command-runner.jar` auslöst. Der in der `Args`-Liste angegebene Befehl lädt ein Skript namens `my-script.sh` von Amazon S3 in das Hadoop-Benutzer-Home-Verzeichnis herunter. Der Befehl ändert dann die Berechtigungen des Skripts und führt `my-script.sh` aus.

Wenn Sie den AWS CLI verwenden, sollten die Elemente in Ihrer `Args` Liste durch Kommas getrennt sein, ohne Leerzeichen zwischen den Listenelementen. Verwenden Sie z. B. `Args=[example-command,example-option,"example option value"]` statt `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 : Ausführen eines Skripts in einem Cluster mit **script-runner.jar**

Wenn Sie `script-runner.jar` verwenden, geben Sie das Skript, das Sie ausführen möchten, in der Argumentliste Ihres Schritts an.

Im folgenden AWS CLI-Beispiel wird ein Schritt an einen aktiven Cluster gesendet, der `script-runner.jar` auslöst. In diesem Fall wird das Skript mit dem Namen `my-script.sh` auf Amazon S3 gespeichert. Sie können auch lokale Skripte angeben, die auf dem Hauptknoten Ihres Clusters gespeichert sind.

```
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=script-runner.jar,Args=[my-script.sh]
```

```
--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]
```

Andere Verwendungsmöglichkeiten von **command-runner.jar**

Sie können `command-runner.jar` auch verwenden, um Arbeiten mit Tools wie `spark-submit` oder `hadoop-streaming` in einen Cluster einzureichen. Wenn Sie eine Anwendung mit `command-runner.jar` starten, geben Sie `CUSTOM_JAR` als Schritttyp an, anstatt einen Wert wie `SPARK`, `STREAMING` oder `PIG` zu verwenden. Die Verfügbarkeit der Tools hängt davon ab, welche Anwendungen Sie auf dem Cluster installiert haben.

Der folgende Beispielbefehl `command-runner.jar` dient zum Senden eines Schritts mit `spark-submit`. In der `Args`-Liste wird `spark-submit` als Befehl angegeben, gefolgt von der Amazon-S3-URI der Spark-Anwendung `my-app.py` mit Argumenten und Werten.

```
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]
```

In der folgenden Tabelle sind zusätzliche Tools aufgeführt, die Sie mit `command-runner.jar` verwenden können.

Name des Tools	Beschreibung
<code>hadoop-streaming</code>	Sendet ein Hadoop-Streaming-Programm. In der Konsole und einige SDKs ist dies ein Streaming-Schritt.
<code>hive-script</code>	Führt ein Hive-Skript aus. In der Konsole und in SDKs ist dies ein Hive-Schritt.
<code>pig-script</code>	Führt ein Pig-Skript aus. In der Konsole und in SDKs ist dies ein Pig-Schritt.
<code>spark-submit</code>	Führt eine Spark-Anwendung aus. In der Konsole ist dies ein Spark-Schritt.

Name des Tools	Beschreibung
hadoop-lzo	Führt die Hadoop-LZO-Indexerstellung in einem Verzeichnis aus.
s3-dist-cp	Verteiltes Kopieren großer Datenmengen aus Amazon S3 in HDFS. Weitere Informationen finden Sie unter S3 DistCp (s3-dist-cp) .

AWS-Glossar

Die neueste AWS-Terminologie finden Sie im [AWS-Glossar](#) in der AWS-Glossar-Referenz.

Die vorliegende Übersetzung wurde maschinell erstellt. Im Falle eines Konflikts oder eines Widerspruchs zwischen dieser übersetzten Fassung und der englischen Fassung (einschließlich infolge von Verzögerungen bei der Übersetzung) ist die englische Fassung maßgeblich.